

投資割合を考慮した投資先類似度に基づく投資信託のクラスタリング

Mutual Funds Clustering Based on Investment Similarity Considering Investment Ratio

米田 一樹 *1
Kazuki Yoneda松井 藤五郎 *2
Tohgoroh Matsui犬塚 信博 *1
Nobuhiro Inuzuka武藤 敦子 *1
Atsuko Mutoh森山 甲一 *1
Koichi Moriyama*1名古屋工業大学
Nagoya Institute of Technology*2中部大学
Chubu University

It is necessary to compare mutual funds with their investment similarity, because their actual operations may differ from their investment policies. A method for clustering mutual funds based on their investment similarity has been proposed, but it only uses top 10 invested companies. In this paper, we propose a new similarity between mutual funds using all invested companies and their investment ratio.

1. はじめに

近年、金融業界ではフィンテック (FinTech) というキーワードで情報技術を活用した金融商品の開発が盛んに行われており、人工知能 (AI) も活用されている。

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。ファンドはそれぞれ運用方針を持っておりそれによって複数の金融商品へ投資を行う。そのため投資家はファンドの性格を把握する必要がある。また、ファンドのパフォーマンスは運用方針により大きく異なるため、運用方針をファンド間で比較する必要がある。

ファンドは運用方針によってパッシブファンドとアクティブファンドの二つに分けられる。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つファンドであり、アクティブファンドはインデックスを超過した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方針が異なる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。

武田らは、「実際の運用方針に基づいたファンドのグループ分け」を目的にファンドに対して投資割合を考慮した投資先類似度を提案し、ファンドをクラスタリングする方法を提案した [武田ら 2013]。また、榊原らは、スペクトラルクラスタリングを用いることで、より有効なクラスタリングができることを示した [榊原ら 2014, Sakakibara et al. 2015]。しかしながら、武田らと榊原らの手法では投資先類似度が投資割合の上位 10 銘柄しか考慮していないため、細かな違いを見ることができなかった。そこで本論文では、投資割合を考慮して全銘柄から類似度を求める方法を提案する。

2. 投資先類似度に基づく投資信託クラスタリング

武田・榊原らは「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、ファンドが公表している

投資割合の高い上位 10 銘柄から投資先類似度を求め、投資先類似度に基づきクラスタリングを行い結果を評価した。

武田・榊原らの手法では、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」とし、ファンド v_i と v_j 間の類似度 $w(v_i, v_j)$ を以下のように定義している。

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

ここで、 v はファンド、 $I(v)$ はファンド v が投資している株式銘柄の集合、 $I_{10}(v)$ はファンド v の投資割合が高い上位 10 の株式銘柄の集合を表す。

榊原らは、スペクトラルクラスタリング [Luxburg 2007] を用いてファンドのクラスタリングを行うことを提案した。スペクトラルクラスタリングは、グラフ分割問題としてクラスタリングを行う手法である。クラスタリングを行う前に、データをノードとし、データ間にその類似度に応じて重みを付けたエッジを作り、グラフを作成する。このグラフを、同じクラスタで構成されたサブグラフは密、異なるクラスタで構成されたサブグラフ間は疎となるようにクラスタリングを行う。このデータ間の類似度として、ファンド間の類似度を用いることで、ファンドのクラスタリングを行う。

3. 提案手法

3.1 従来手法の問題点

ファンドはそれぞれ運用方針に従って投資を行っている。ファンドの中には結果的に運用方針と実際の運用方針が一致せず、方針と結果が乖離してしまう場合がある。そのため投資家にとっては公表されている運用方針よりも実際の運用方針が重要である。投資家はファンドを選ぶ際、ファンドが一つの株式銘柄にどれだけ投資したかを示す投資割合を考慮してファンドを選ぶ。

また、投資先上位 10 社目までの情報を用いてクラスタリングを行ったが、投資先上位 11 社目以降を考慮していない。例えば投資先上位 10 社目の株式への投資割合が 1.56% で 11 社目の投資割合が 1.55% の場合、近い投資割合で投資しているが 11 社目は考慮されない。しかし実際の投資においては 11 社目以降も考慮して投資先の決定が行われる。

そこで本研究では榊原が提案した投資先類似度に全投資先の投資割合を考慮した新たな投資先類似度を提案する。

連絡先: 米田一樹, 名古屋工業大学情報工学科犬塚研究室, 名古屋市昭和区御器所町, 052-735-5050, k.yoneda.690@nitech.jp

表1 ファンド v_1 の投資割合上位 20 銘柄

順位	銘柄コード	銘柄名	組み入れ比率
1	9983	ファーストリテイリング	6.420%
2	9433	KDDI	4.354%
3	6954	ファナック	3.673%
4	6971	京セラ	2.509%
5	6367	ダイキン工業	2.106%
6	4543	テルモ	2.025%
7	9735	セコム	1.967%
8	8035	東京エレクトロン	1.815%
9	4503	アステラス製薬	1.700%
10	4523	エーザイ	1.558%
11	5563	日東電工	1.546%
12	4063	信越化学工業	1.436%
13	6762	TDK	1.421%
14	4507	塩野義製薬	1.399%
15	4452	花王	1.364%
16	7267	本田技研工業	1.356%
17	9613	NTT データ	1.297%
18	7203	トヨタ自動車	1.264%
19	4324	電通	1.244%
20	4502	武田薬品工業	1.086%

表2 ファンド v_2 の投資割合上位 20 銘柄

順位	銘柄コード	銘柄名	投資割合
1	9983	ファーストリテイリング	7.824%
2	9984	ソフトバンク	4.077%
3	6954	ファナック	4.023%
4	9433	KDDI	3.855%
5	6971	京セラ	2.189%
6	6367	ダイキン工業	2.147%
7	8035	東京エレクトロン	2.048%
8	4543	テルモ	1.736%
9	4063	信越化学工業	1.690%
10	9735	セコム	1.638%
11	4503	アステラス製薬	1.583%
12	5563	日東電工	1.546%
13	6762	TDK	1.505%
14	8028	ユニーファミリーマート	1.431%
15	4523	エーザイ	1.374%
16	7267	本田技研工業	1.256%
17	7203	トヨタ自動車	1.241%
18	9613	NTT データ	1.198%
19	4452	花王	1.128%
20	2432	DeNA	1.095%

3.2 投資割合を考慮した類似度

ファンド v が投資している株式銘柄の集合を $I(v)$ 、上場されている株式銘柄の集合を A 、その要素を $a \in A$ とする。ファンド v が株式銘柄 $a \in A$ への投資割合を $t(v, a)$ と表す。

このとき、ファンド v_i と v_j 間の類似度 $T(v_i, v_j)$ を以下のよう

$$T(v_i, v_j) = \sum_{a \in A} \min\{t(v_i, a), t(v_j, a)\}$$

すなわち、全ての上場銘柄について、その銘柄への投資割合を比較し、その小さい方の総和を類似度とする。ファンド v_i と v_j の類似度が $T(v_i, v_j) = 0.5$ のとき、 v_i と v_j の総資産額が等しいとすると、総資産額の半分が同じ銘柄で構成されていることを意味する。

4. 実験と考察

4.1 実験方法

提案した類似度の有効性を確認するために、三菱 UFJ 国際投信株式会社が運用している下記の 5 つのファンドについて、2017 年 3 月 6 日の時点で公開されていた運用報告書（全体版）を Web サイトよりダウンロードし、そこに掲載されていた組み入れ資産の明細に基づいて従来手法の類似度と提案手法におけるファンドの類似度を求め、比較した。

1. 〈DC〉インデックスファンド 225
2. 三菱 UFJ インデックス 225 オープン
3. eMAXIS 日経 225 インデックス
4. いちよし 公開ベンチャー・ファンド
5. 三菱 UFJ 技術のチカラ オープン

以後、それぞれ、ファンド v_1, \dots, v_5 と表す。

v_1 から v_3 は、日経平均株価（日経 225）と連動する投資効果を目指しているパッシブファンドである。 v_4 と v_5 は、成長性が高く、かつ、株価水準が割安であると判断される銘柄を中心に投資するアクティブファンドである。表 1-5 に、各ファンドの投資割合上位 20 銘柄を示す。ここで、 v_1 と v_3 は「日経 225

表3 ファンド v_3 の投資割合上位 20 銘柄

順位	銘柄コード	銘柄名	投資割合
1	9983	ファーストリテイリング	8.450%
2	6954	ファナック	4.127%
3	9433	KDDI	3.791%
4	9984	ソフトバンク	3.323%
5	6971	京セラ	2.247%
6	4503	アステラス製薬	1.822%
7	9735	セコム	1.777%
8	6367	ダイキン工業	1.692%
9	4523	エーザイ	1.622%
10	4543	テルモ	1.620%
11	5563	日東電工	1.617%
12	8035	東京エレクトロン	1.577%
13	6762	TDK	1.521%
14	7203	トヨタ自動車	1.514%
15	7267	本田技研工業	1.481%
16	4452	花王	1.361%
17	4324	電通	1.359%
18	4063	信越化学工業	1.338%
19	4502	武田薬品工業	1.279%
20	9613	NTT データ	1.259%

マザーファンド」をマザーファンド *1 とするベビーファンド *2 であり、 v_2 は「三菱 UFJ インデックス 225 マザーファンド」をマザーファンドとするベビーファンドである。

4.2 実験結果

従来手法と提案手法におけるファンド間の類似度を表 6 と表 7 に示す。ここで、比較のため、従来手法の類似度は最大で 1 となるように 10 で割って正規化してある。

従来手法では、 v_1 と v_3 の類似度が 0.9 と最も高く、次いで v_1 と v_3 、 v_2 と v_3 の類似度が 0.8 と高くなった。また、アクティブファンド群 (v_4, v_5) とパッシブファンド群 (v_1, v_2, v_3) の間の類似度は 0、アクティブファンド間の類似度は 0.1 であった。

提案手法でも、 v_1 と v_3 の類似度が 0.8784 と最も高く、次い

*1 自らに投資するベビーファンドの資金をまとめることによって、効率を高めて運用するファンド。

*2 マザーファンドに投資するファンド。実質的な運用はマザーファンドで行う。

表4 ファンド v_4 の投資割合上位 20 銘柄

順位	銘柄コード	銘柄名	投資割合
1	2484	夢の街創造委員会	5.285%
2	4584	ペプチドリーム	4.096%
3	3679	じげん	3.664%
4	6058	ベクトル	3.145%
5	7779	サイバーダイナ	2.765%
6	6099	エラン	2.574%
7	7636	ハンズマン	2.534%
8	7747	朝日インテック	2.404%
9	6145	日特エンジニアリング	2.297%
10	4293	セブテーニ	2.281%
11	5975	東プレ	2.247%
12	6240	ヤマシンフィルタ	1.920%
13	7743	シード	1.891%
14	7613	シークス	1.886%
15	2301	学情	1.821%
16	8715	アニコムホールディングス	1.779%
17	6312	フロイント産業	1.614%
18	4186	東京応化工業	1.547%
19	7867	タカラトミー	1.515%
20	6071	IBJ	1.505%

表5 ファンド v_5 の投資割合上位 20 銘柄

順位	銘柄コード	銘柄名	投資割合
1	7408	ジャムコ	3.817%
2	6005	三浦工業	3.469%
3	7747	朝日インテック	2.818%
4	6826	本多通信工業	2.704%
5	7701	島津製作所	2.656%
6	6289	技研製作所	2.517%
7	7600	日本MDM	2.464%
8	6951	日本電子	2.291%
9	6013	タクマ	2.046%
10	6345	アイチコーポレーション	2.042%
11	6859	エスベック	1.999%
12	6810	日立マクセル	1.944%
13	4584	ペプチドリーム	1.943%
14	6622	ダイヘン	1.859%
15	6920	レーザータック	1.833%
16	6312	フロイント産業	1.800%
17	2009	鳥越製粉	1.779%
18	4996	クミアイ化学工業	1.702%
19	2915	ケンコーマヨネーズ	1.635%
20	1377	サカタのタネ	1.608%

表6 従来手法における類似度

	v_2	v_3	v_4	v_5
v_1	0.8	0.9	0.0	0.0
v_2		0.8	0.0	0.0
v_3			0.0	0.0
v_4				0.1

で v_1 と v_3 の類似度が 0.8734、 v_2 と v_3 の類似度が 0.8717 と高くなった。 v_4 とパッシブファンド群の間の類似度は 0、 v_5 とパッシブファンド群の間の類似度は 0.0065 から 0.0067 であった。アクティブファンド間の類似度は 0.0833 であった。

4.3 考察

アクティブファンド群 (v_4, v_5) とパッシブファンド群 (v_1, v_2, v_3) との類似度について見ると、従来手法では類似度は全て 0 であったが、提案手法では v_4 の類似度は 0、 v_5 の類似度は 0 にはならなかった。従来手法では、類似度の計算に組み入れ比率

表7 提案手法における類似度

	v_2	v_3	v_4	v_5
v_1	0.8434	0.8784	0.0000	0.0065
v_2		0.8718	0.0000	0.0067
v_3			0.0000	0.0066
v_4				0.0833

上位 10 社しか用いていないため、類似度が 0 であっても同じ銘柄に投資している可能性がある。これに対し、提案手法では、全銘柄を用いて類似度を計算しているため、類似度が 0 であるときは同じ銘柄に投資していないことがわかる。つまり、提案手法を用いることによって、 v_4 の組み入れ銘柄はパッシブファンド群と一つも一致していないが、 v_5 の組み入れ銘柄の中にはパッシブファンド群の組み入れ銘柄と同じものがわずかに含まれていることがわかる。

v_1 と v_3 はマザーファンドが同じであるため、組み入れ比率はほぼ同じになっているはずである。従来手法、提案手法共に類似度が 1 にならないのは、ファンドの決算日が v_1 が 2016 年 5 月 23 日、 v_3 が 2016 年 1 月 26 日と異なり、開示されている情報の基準日が異なるためである。2016 年 4 月 1 日に、日経平均株価の構成銘柄から横浜銀行が除外され、コンコルディア・フィナンシャルグループが採用された。また、 v_2 に組み入れられているユニーフAMILYマートは、ユニーフAMILYマートの合併によつた新たに誕生した銘柄であり、合併前はユニーフAMILYが組み入れられている。このように、基準日が異なることによって組み入れ比率だけでなく構成銘柄にも影響があるため、同じ基準日の組み入れ比率を用いて類似度を計算することが望ましいと考えられる。

5. まとめ

本論文では、投資信託クラスタリングにおける類似度尺度として、投資割合を考慮した投資先類似度を提案した。従来手法では組み入れ比率上位 10 銘柄のみを用い、その共通銘柄数を類似度としていたが、11 位以下の銘柄が考慮されていないため、細かな違いを見ることができなかった。提案手法では、全銘柄を用いることで、細かな違いを見ることができるようになった。

本論文では公開されている運用報告書に記載されている組み入れ比率を用いて類似度を求めたが、基準日が異なると類似度を正確に求めることができないことが明らかになった。

今後は、提案した類似度を用いてクラスタリングを行い、従来手法に対する提案手法の有効性を確認したい。

参考文献

- [武田ら 2013] 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)
- [神原ら 2014] 神原拓将, 松井藤五郎, 犬塚信博, 武藤敦子: スペクトラルクラスタリングを用いた投資信託分析, 第 28 回人工知能学会全国大会 2014 (JSAI2014), 3L3-OS-26a-5 (2014)
- [Sakakibara et al. 2015] Sakakibara, T., Matsui, T., Mutoh, A., Inuzuka, N.: "Clustering mutual funds based on investment similarity", *Procedia Comp. Sci.*, 60:881-890, KES2015 (2015).
- [Luxburg 2007] U. von Luxburg: "A tutorial on spectral clustering," *Stat. Comp.*, 17(4):395-416 (2007)