

因果関係分析に基づく投資関連商品の比較手法の検討

栗納 裕貴[†] 馬 強[†] 吉川 正俊[†]

[†] 京都大学情報学研究科 社会情報学専攻 分散情報システム分野 〒 606-8501 京都市左京区吉田本町

E-mail: [†]awano@db.soc.i.kyoto-u.ac.jp, ^{††}{qiang,yoshikawa}@i.kyoto-u.ac.jp

あらまし NISA(少額投資非課税制度)の導入など一般人が投資信託など投資関連商品を比較検討し購入する機会が増加している。しかし、投資関連商品は非常に多種におよび、また、商品内容は日常生活となじみがうすいため目論見書を読んだのみでは商品の理解が困難であり、各商品の特徴を理解して商品を選択することは困難である。本研究では投資関連商品と現実世界での経済事象の因果関係分析を行い、因果関係図をユーザーへ提示することで商品について現実世界との関係を踏まえ理解することを支援し、また、複数の商品を一つの因果関係図の中に含め提示することで、複数商品の中で各商品の特徴の比較・理解を容易に行えるようにする手法を検討する。

キーワード 因果関係, 投資, 比較

1. はじめに

近年 NISA の導入に加え株高の背景があり、一般人が投資信託など投資関連商品を比較検討し購入する機会が増加している。NISA での口座開設数は 2014 年 1 月 3 日付けの日本経済新聞社での報道によれば開設見込みをあわせて 400 万件^(注1)を超えている。NISA は少額投資に対する減税措置であるため、少額で購入でき、なおかつ株式などに比べ分散投資が行えリスクが抑えられる投資信託を選択する個人投資家が増えることが期待される。

その一方で投資の対象となる投資関連商品はその商品の種類が非常に多種に及ぶ。先の日本経済新聞社の記事によれば、証券会社や銀行が用意する NISA 向けの投資信託商品のみでも、600 本以上に及ぶ。また、投資関連商品はその商品内容、投資先が株式や不動産など普段の生活と馴染みがうすく、商品の目論見書などを読んだのみでは商品について理解をすることは困難であることがある。

投資信託では通常月次レポート、または商品によっては週次レポートにより商品の投資の現状及び今後についての報告が行われる。また、通常レポートの中ではファンドマネージャーからのコメントなどとして、社会事象を踏まえた投資の成績の解説が行われる。しかし、これらの月次レポートは毎月発行されるため、そのレポートの数は多い。例えば過去 1 年について月次レポートを読み、ある商品についてその商品が現実の社会事象とどういった関係にあるのかを把握しようとする 12 本のレポートを読む必要がある。ここで例えば、現在、検討中の商品が 10 ある場合、それぞれの商品にレポートが存在するため、120 本のレポートを読むこととなり、個人投資家にとっては大きな負担である。

投資信託に関するアンケート調査報告書-2012 年^(注2)による

と、投資信託保有未経験者の非購入理由では 55.1% で「投資信託がよく分からないので不安」が最大の理由としてあげられている。また、保有未経験者層全体では「商品の種類が多く自分で選択できない」を非購入理由としてあげる割合は 16.2% であるが、保有未経験者の中の購入意向層ではこの割合は 26.3% となり、購入を行いたい商品の商品の種類が多く選択が行えない問題が存在していることが伺える。

既存のサービスにおいて投資信託の商品のランキングや投資信託の比較が行えるサービスがある。しかし、それらは人気ランキング^(注3)であったり、または基準価額などの商品の属性情報やチャートの比較^(注4)であり、投資商品の投資内容の特徴による比較ではない。

長期に渡る投資においては、投資信託商品の特徴を十分に理解し、投資を行うことが重要である。長尾 真 著「わかる」とは何か [1] では「わかる」には 3 つのレベルがあるとしている。

第 1 のレベル 言葉の範囲内で理解する

第 2 のレベル 文が述べている対象世界との関係で理解する

第 3 のレベル 自分の知識と経験、感覚に照らして理解する

第 1 のレベルである言葉の範囲内で理解するとは、例えば J-REIT の商品について J-REIT とは国内不動産投資であるということを辞書的に理解したレベルである。これは既存の検索エンジンなどを用い、辞書的に意味が分からない単語を調べていくことにより達成できると考えている。

第 2 のレベルである文が述べている対象世界との関係で理解するとは、先の例で言えば J-REIT の価格が上昇するのは国内の景気が上向いた場合であり、また、国内の景気は東京オリンピックなどによる国内での不動産需要に影響されることまで理解したレベルである。本研究ではこの第 2 のレベルの理解を目標とする。

本論文では投資信託の過去の月次レポートを自然言語処理技術を用いて解析し、複数商品についてそれらの商品と現実の経

(注1): N I S A口座数、400 万超す 高まる個人の投資意欲 [http://www.nikkei.com/article/DGXNASGD2706E_\\\$4A100C1MM8000/?dg=1](http://www.nikkei.com/article/DGXNASGD2706E_\$4A100C1MM8000/?dg=1)

(注2): 投資信託に関するアンケート調査報告書-2012 年 <http://www.toushin.or.jp/statistics/report/research2012/>

(注3): 価格.com 投資信託比較 <http://kakaku.com/fund/>

(注4): モーニングスター <http://www.morningstar.co.jp/>

済事象との間の関係の因果関係ネットワークを構築しユーザーへ提示することで、ユーザーがそれぞれの商品と経済事象の関係及び、それぞれの商品の特性を理解し、投資先を決定することを支援することを目的とする。

本論文におけるシステムの最終出力のイメージを図1に示す。また、本論文における因果関係の特徴として、各エッジが次の特徴を持つことがある。

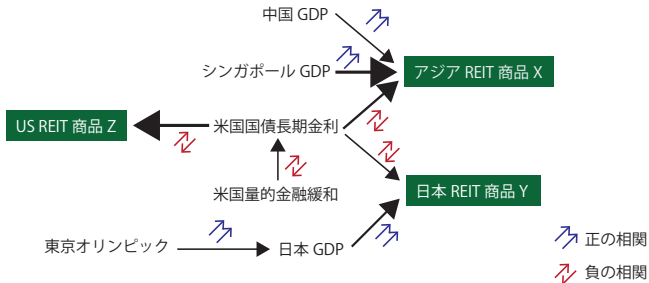


図1 因果関係に基づく投資商品の比較イメージ

関係の強さ 各因果関係の影響がエッジには割り当てられている。図1ではアジア REIT においては中国 GDP よりもシンガポール GDP が強い影響を持っている。

関係の極性 各因果関係の事象間の関係が正の相関であるか、負の相関であるかが割り当てられている。図1では米国国債長期金利が低下すると日本 REIT の基準価格は上昇する関係であることが分かる。

通常、投資信託では複数の投資先から商品が組まれるため、1つの投資信託の商品に対して複数の要因が存在することが一般的である。従って、それぞれの要因が投資信託の商品に対して、どの程度の影響を与えるのか関係の強さを知ることは重要である。また、経済事象においては相関関係を持った関係が存在する。以降において原因が A であり結果が B である因果関係について $A \rightarrow B$ と記述する。例えば、米国国債長期金利が低下すると US REIT の基準価額は上昇する一方、米国国債長期金利が上昇すると US REIT の基準価額は低下する負の相関は、

(1) 米国国債長期金利が低下 $\rightarrow$ USREIT の基準価額の上昇

(2) 米国国債長期金利が上昇 $\rightarrow$ USREIT の基準価額の低下

の2つの因果関係として提示するより、

米国国債長期金利 $\rightarrow$ 負の相関 USREIT の基準価額

の1つの因果関係として提示する方がより容易に理解することができると考える。本論文において因果関係の抽出は、商品の月次レポート内の記述及び、レポートが対象とする期間の商品の基準価額の変動を用いて行う。図2に月次レポート内の記述の例を示す。

月次レポート内の記述からの因果関係の抽出には手がかり表現を用いる。図2では、「しかし、後半には米国の量的金融緩和の縮小が先送りされるとの観測から長期金利が低下したため、市場は緩やかに上昇する展開となりました。」において手がかり表現「から」より「米国の量的金融緩和の縮小が先送りされ

る」を要因として「長期金利が低下」の結果がもたらされた因果関係を得る。

米国の量的金融緩和の縮小が先送りされる $\rightarrow$ 長期金利が低下

次にレポートが対象とする期間の基準価額の変動を用いて因果関係を抽出する。図2では、前月末から289円の上昇があったことが記されている。従って、このレポート内に記述されている因果関係はこの商品に対して基準価額の上昇の結果を及ぼした可能性が高い。そこで、先に手がかり表現を用いて取得した因果関係は基準価額の上昇の結果を及ぼしたと推測し、次の因果関係を得る。

米国の量的金融緩和の縮小が先送りされる $\rightarrow$ 長期金利が低下 $\rightarrow$ アジア REIT・リサーチ・オープンの基準価額の上昇

また、同商品に関する作成基準日2013年11月29日のレポートより次の因果関係を得る。

米国で量的金融緩和の縮小の開始時期が早まる $\rightarrow$ 長期金利が上昇 $\rightarrow$ アジア REIT・リサーチ・オープンの基準価額の低下

次に、事象の極性を求める。ここで、事象について本論文においては次の形式で表現を行う。事象 A の記述に用いられている名詞節の集合を W_n 、動詞節の集合を W_v 、事象 A の極性を $p = \{+or-\}$ とした場合、 $A = (W_n, W_v, p)$ の3つ組で表現を行う。ここで節とは1つの以上の単語が連結されたもので、名詞節とは「量的金融緩和」など名詞的な役割を持つ節、また動詞節とは「先送り」や「縮小」など動詞的な役割を持つ節である。

先の例をこの表現を用いて示すと以下になる。

(1) ($\{ \text{米国, 量的金融緩和} \}, \{ \text{縮小, 先送り} \}, + \rangle \rightarrow (\{ \text{長期金利} \}, \{ \text{低下} \}, -) \rightarrow (\{ \text{アジア REIT・リサーチ・オープン, 基準価額} \}, \{ \text{上昇} \}, +)$

(2) ($\{ \text{米国, 量的金融緩和, 開始時期} \}, \{ \text{縮小, 早まる} \}, -) \rightarrow (\{ \text{長期金利} \}, \{ \text{上昇} \}, +) \rightarrow (\{ \text{アジア REIT・リサーチ・オープン, 基準価額} \}, \{ \text{低下} \}, -)$

ここで説明のため、矢印に下付けで相関の正負を記述する。矢印の開始点と終了点での極性が同じであれば正を割り当て、異なっていれば負を割り当てる。

(1) ($\{ \text{米国, 量的金融緩和} \}, \{ \text{縮小, 先送り} \}, + \rangle \rightarrow_{\text{負}} (\{ \text{長期金利} \}, \{ \text{低下} \}, -) \rightarrow_{\text{負}} (\{ \text{アジア REIT・リサーチ・オープン, 基準価額} \}, \{ \text{上昇} \}, +)$

(2) ($\{ \text{米国, 量的金融緩和, 開始時期} \}, \{ \text{縮小, 早まる} \}, -) \rightarrow_{\text{負}} (\{ \text{長期金利} \}, \{ \text{上昇} \}, +) \rightarrow_{\text{負}} (\{ \text{アジア REIT・リサーチ・オープン, 基準価額} \}, \{ \text{低下} \}, -)$

これらの因果関係に対して事象の同一性判定と統合を行い、同一事象をまとめ1つの事象に統合することで次の因果関係を最終的に得る。

米国の量的金融緩和 $\rightarrow_{\text{負}}$ 長期金利 $\rightarrow_{\text{負}}$ アジア REIT・リサーチ・オープン

本論文では因果関係の強さについても考慮をする。関係の強さは確信度、支持度、及び影響度(影響の大きさ)より求める。確信度、及び支持度は求められた因果関係の出現回数より求める。影響度は、その因果関係が抽出された月次レポート

【市場動向の振り返り】

10月のアジアREIT市場は、香港が横ばい、シンガポールが小幅上昇となりました。前半は、米国の連邦政府機関の一部閉鎖や連邦債務上限の引き上げを巡る議会の紛糾などを受けて投資家心理が悪化、市場は軟調に推移しました。しかし、後半には米国の量的金融緩和の縮小が先送りされるとの観測から長期金利が低下したため、市場は緩やかに上昇する展開となりました。月半ばよりシンガポールのREIT企業の決算発表が始まり、市場予想を上回る堅調な業績を発表した銘柄の株価が上昇する場面がありました。月間では HANG SENG REIT 指数が+0.1%、FTSE ST REIT 指数が+1.3%となりました。
(市場変化率は、9/27～10/30の期間にて算出しています。)

図 2 アジア REIT・リサーチ・オープン (毎月決算型) 三井住友トラスト・アセットマネジメント 作成基準日 2013 年 10 月 31 日

における基準価額の変動より割り当てる。これは基準価額が大きく変動した月の月次レポートにおける因果関係は基準価額の変動に対して大きな影響を持っていた可能性が高いことを仮定している。

因果関係の抽出を複数の投資関連商品に対して行い、複数の投資関連商品と対象世界との関係をひとつの因果関係図として示す。これにより複数の投資関連商品と対象世界との関係の理解を支援し、それぞれの商品の投資内容の特徴を理解し商品の選択を行えるようにする。

本研究における技術的課題を以下に示す。

(1) **因果関係の記述の抽出** 自然言語で記述された月次レポートから因果関係を抽出する課題がある。図 2 の「投資家心理が悪化、市場は軟調に推移」にあるように明示的に手がかり表現が出現しなかったり、「アジア REIT 市場」がレポート内では省略され「市場」として記述されているなど月次レポート特有の課題が存在する。

(2) **事象の表現モデル** 同一性判定を行い因果関係図を構築するために、事象を適切にモデル化する必要がある。本研究では極性及び関係の相関を考慮した因果関係図の作成を目標とするため、事象の極性に関する情報を分割して保持する必要がある課題が存在する。

(3) **事象の同一性判定** 例えば量的金融緩和、QE3 など、同じ事象について異なる記述が用いられることがあり、これらの同一性判定の課題がある。また、これらの語は日常語でなく経済用語などの専門用語であり、新しい用語が次々に登場することが課題として存在する。

2. 関連研究

2.1 手がかり表現を利用した因果関係の抽出

手がかり表現を用いた因果関係の抽出では、坂地 [2] らは構文パターンと手がかり表現を用いて因果関係を抽出する研究を行なっている。日本経済新聞より得た景気動向に関する 300 の記事から、手がかり表現の取得を行なっている。その後、4 つの構文パターンを用いて因果関係に関する知識抽出を行う手法を提案している。

2.2 事象の類似性判定

事象の類似性判定については、石井 [3] らが構造化したキーワードと事象のトピックを用いた事象の表現モデル及び、トピックを考慮した因果関係ネットワークの増分的な構築の手法を提案している。提案手法においては事象の類似性判定をトピックごとに行うことで事象の類似性判定に必要な計算量の削減を行なっている。

また、キーワードの比較の語の概念距離計算の手法を複数導入しそれらにより事象ノード対の SVO 組の関連度計算を行い類似性を判定している。その際、元の記事での文脈を利用しノードに含まれる各キーワードが使われている意味を特定している。

2.3 事象の極性判定

河合 [4] らは、対応する名詞句についての変化、行動、動作や動きを表す句である TVP(傾向動詞句, Trend Verbal Phrase) の概念を提案し、事象及びその関係の極性を求め、因果関係ネットワークを求める手法を提案した。

3. 提案手法

3.1 提案手法の概略

提案手法では入力として複数の投資信託の複数の月次レポートを受け取る。それらを入力とし、複数の投資信託の商品と経済事象の間における因果関係のネットワークを構築し出力する。因果関係の取得については次の手順で行う。

(1) 月次レポート内の記述から手がかり表現を利用して因果関係を抽出 (§3.2) し、抽出された因果関係に含まれる事象について事象の同一判定と統合 (§3.5) を行う。これにより取得した因果関係の集合を U とする。

(2) 次に基準価額を利用して因果関係を獲得 (§3.6) し、それらを U へ追加する。

(3) U の因果関係について確信度、支持度 (§3.7) 及び影響度 (§3.8) を求める。

(4) 取得された因果関係を元にネットワークを構築し、ユーザへ提示する。

3.2 手がかり表現を利用した因果関係の抽出

月次レポート内の記述を構文木解析し、手がかり表現により因果関係を抽出する。発見された手がかり表現にかかる文節を構文木を遡っていく方向で取得していき、それらを連結したものを原因記述 (C_{exp}) とする。発見された手がかり表現がかかる文節を構文木を下っていく方向で文末または動詞を含む文節まで取得していき、それらを連結したものを結果記述 (R_{exp}) とする。また、結果記述を取得する際については、動詞を含む文節があった場合は、その文節から構文木を遡っていく方向で取得し、それらを結果記述に加える。これは動詞の主部を取得するためである。

3.3 名詞節及び動詞節の獲得

名詞節及び動詞節については、次の手順を 2 度実行することにより獲得する。1 回目の実行では、用語リストは空として手

順を実行し獲得した名詞節を用語リストに追加し、用語リストを構築する。2回目の実行では、1回目の実行により獲得した用語リストを利用し名詞節、及び動詞節を獲得する。

(1) 構文木解析により取得したチャンク内の単語をすべて連結する。ただし、チャンクの末尾が記号、助詞、助動詞で終わっている場合は、それらの単語は削除する。

(2) 動詞で終わっている場合は動詞節とする。それ以外の場合は名詞節とする。

(3) 単一の名詞しか含んでなく、また、その名詞がサ変接続名詞である場合は動詞節とする。

(4) 用語リストに登録されている名詞節の後に、サ変接続名詞が続く節が終わる場合、用語リストに含まれている名詞節の部分を名詞節とし、サ変接続名詞の部分を動詞節とする。

用語リストを利用することで、名詞節にサ変接続名詞が接続された記述を適切に分割することができる。例えば、「量的金融緩和縮小」の記述は量的金融緩和、縮小の2つの節に別けられるべきである。この場合、量的金融緩和が用語リストに登録されていることで、量的金融緩和と縮小の2つに分けることができる。また、この手順により獲得した用語リストの例を表1に示す。

表1 獲得した用語の例

年金積立金管理運用独立行政法人、大型商業 REIT、投資家心理、公的・準公的年金、地価 LOOK レポート、高度利用地価動向報告、欧米株式市場、量的金融緩和、金融政策決定会合、内外株、...

3.4 事象の極性判定

事象の記述に含まれる動詞節の集合を用いて、事象の極性を判定する。動詞節について予め極性を割り当てた辞書を用意しておき、それを利用し判定を行う。この辞書について、既存のものが利用できるかは検討中である。極性について、大きさを持つもの場合は大きくなる方向が+側であり、また時間の概念を持つものについては早くなる方向が+側である。表2は極性を動詞節に割り当てた例である。

表2 動詞節の極性の例

極性	動詞節
+	拡大, 早まる, ある
-	縮小, 先送り, ない

ここで、単語集合の中で-の極性を持つ単語の数を2で割った場合の余りが0である場合、その事象の極性は+であるとす。それ以外については事象の極性として-を与える。また、事象の記述に動詞節が含まれなかった場合は、事象の極性として+を与える。これは事象の記述がある時点において、その事象があることが影響を及ぼしたことを記述者は記述しようとしていることを仮定している。

動詞節の集合が{縮小,先送り}の例においては、それぞれの単語は-の単語であるため、-の単語の数は2であり+を割り当てる。動詞節の集合が{縮小,早まる}の例においては、

縮小が-であり早まるが+であるため、-の単語の数は1であり-を割り当てる。

3.5 事象の同一性判定及び統合

事象の同一性判定はトピックモデルである LDA により構築したトピックを利用して行う。Blei ら[5]は文章はトピックにより構成され、そしてトピックは単語により構成されるというトピックモデルを提案した。トピック t_i はトピックにおける単語 w_j の出現確率 $P(w_j|t_i)$ により構成され、また、文書 d_k は文書におけるトピックの出現確率 $P(t_i|d_k)$ により構成される。

トピックを特徴量として利用することで、量的金融緩和、量的金融緩和政策、QE3 などの表記揺れを考慮した事象の同一性判定を行うことを期待している。事象に含まれる名詞節を単語とし、事象を文書と見て LDA を利用しトピックを構築する。

事象の同一性判定については、事象 I のトピック t_i に対する確率 $P(t_i|I)$ のベクトルのコサイン類似度に基づき行う。コサイン類似度が閾値以上の場合、事象の統合を行う。

ここで $C(A, B)$ は事象 A, B の各事象のトピックに対する確率のベクトルのコサイン類似度、事象の集合を U 、事象を統合する閾値を σ とすると、事象の同一性判定及び統合は以下の手順で行う。 $C(A, B)$ を最大にする事象 $A, B(\arg\max(C(A, B))s.t. A, B \in U)$ を求める。ここで $C(A, B) \leq \sigma$ であれば、計算終了である。 $C(A, B) > \sigma$ であれば、事象 A 及び B を統合し、計算を続ける。

$N = |U|$ とした時、初めにすべての $C(A, B)$ を計算するのに N^2 が必要であり、その後、統合が起きると N が更新に必要である。最大で N 回の統合が起きるため、事象の同一性判定及び統合における計算量は $O(N^2)$ である。

3.6 基準価額を利用した因果関係の獲得

投資信託の月次レポートでは自然言語の記述以外に基準価額などの数値に関する記述がある。投資信託の基準価額を利用することで、レポート内に記述されている経済事象と投資信託の商品の因果関係の抽出を行う。

商品 x についての月次レポートがあったとする。その月次レポートにおける月末の基準価額を P_e とし、その先月末の基準価額を P_s とする。ここで月次レポートより抽出された因果関係が $A \rightarrow B$ であった場合、次の因果関係を新しく得る。

$$\begin{cases} B \rightarrow (\{x, \text{基準価額}\}, \{\text{上昇}\}, +) & (P_e - P_s \geq 0) \\ B \rightarrow (\{x, \text{基準価額}\}, \{\text{低下}\}, -) & (P_e - P_s < 0) \end{cases} \quad (1)$$

3.7 確信度及び支持度の計算

因果関係のノード間のエッジにおける確信度及び支持度を計算する。確信度及び支持度については、それぞれ式 2,3 により求めることができる。ここで $f(A \rightarrow B)$ は因果関係 $A \rightarrow B$ が抽出された回数である。

$$\text{支持度}(A \rightarrow B) = \frac{f(A \rightarrow B)}{\sum_{A' \rightarrow B' \in U} f(A' \rightarrow B')} \quad (2)$$

$$\text{確信度 } (A \rightarrow B) = \frac{f(A \rightarrow B)}{\sum_{A \rightarrow B' \in U} f(A \rightarrow B')} \quad (3)$$

3.8 影響度の計算

月次レポートにおける、その月末の基準価額を P_e 、その先月末の基準価額を P_s とした場合、影響度 (i) は $i = P_e - P_s$ で与える。

4. 予備実験

手がかり表現を用いた因果関係の抽出、LDA を利用した事象の同一性判定及び統合、獲得した因果関係の可視化に関する予備実験を行った。

構文木解析には日本語の構文木解析のソフトウェア *Cabocha*^(注5) を利用した。構文木解析を行った後、表 3 の手がかり表現を利用し §3.2 より因果関係の取得を行った。また、表 4 を用いて事象の極性の判定を行った。LDA のトピック数は 9 を利用した。また、4.3.3 にて考察を行なっているが各事象におけるトピックに対する確率が平均的であったため、事象間のコサイン類似度を直接計算すると非常に高い類似度となったため、1/トピック数 未満の確率 $P(t_i|I)$ に関しては確率を 0 としコサイン類似度を計算した。また、事象を統合する閾値については 0.8 を利用した。

表 3 利用した手がかり表現

背景に、挙げる、ため、伴う、伴い、ために、により、受けて、効果、影響も、影響が、によって、から

表 4 動詞節の極性

正	下で負に掲げる動詞節以外すべての動詞節
負	崩壊、圧縮、緊縮、カット、減少、縮小、先送り、低下、閉鎖、削減、鈍化、ない、減る

4.1 データセット

表 5 内の商品の月次レポートから取得した記述を予備実験におけるデータセットとして利用した。これらのレポートに含まれる文の数は 101 文であり、84 の因果関係が含まれている。

4.2 結果

プログラムを実行し得た結果について人間により抽出された因果関係が適切かどうかを判断し、表 6 の評価を得た。再現率及び適合率については式 4、5 の通りである。また、LDA により構築されたトピックを表 7 に示す。

表 6 実験結果

再現率	0.27
適合率	0.64

$$\text{再現率} = \frac{|\text{抽出された因果関係のうち適切なもの}|}{|\text{データセットに存在する因果関係}|} \quad (4)$$

$$\text{適合率} = \frac{|\text{抽出された因果関係のうち適切なもの}|}{|\text{抽出された因果関係}|} \quad (5)$$

4.3 考察

4.3.1 因果関係の抽出

抽出された因果関係の例を表 8 に示す。0 の事例において「量的金融緩和縮小」の記述を名詞節「量的金融緩和」と動詞節「縮小」に分け獲得することができている。これより §3.3 の手法が有効であったと言える。一方、5 の事例においては原因の事象が ((根強い), (), +) になっており、事象が記述「高水準の配当利回りを享受できる投資対象としての魅力は根強い」を適切に表現できておらず、適切に事象を抽出できていない。図 3 に *Cabocha* により取得したこの記述の構文木を示す。構文木より手がかり表現「から」にかかる節が「根強い」のみになっていることが適切に事象を抽出できていない原因であることが分かる。

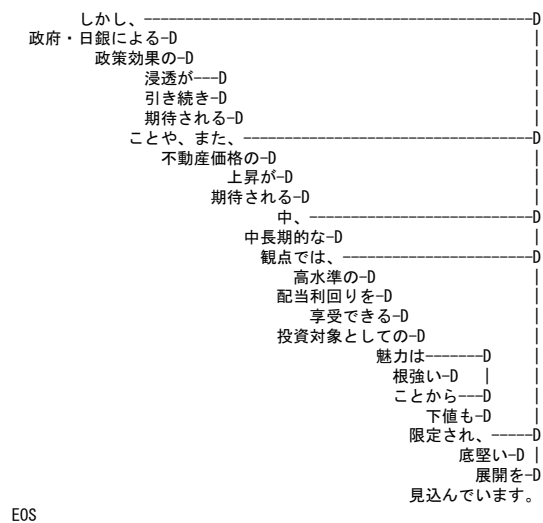


図 3 Cabocha により獲得された構文木

4.3.2 極性の判定

表 8 における事象 1 の原因事象 ((米国, 量的金融緩和, 開始時期), (縮小, 早まる), -) において、「縮小」、「早まる」から事象の極性 - を適切に判定できている。一方で事象 3 の原因事象及 ((米国, 連邦政府機関, 一部), (縮小, 先送りされる), -) 及び結果事象 ((長期金利), (低下し), +) において、それぞれ +, - であるべきであるが、適切に判定ができていない。「先送り」及び「低下」は極性の判定のための辞書に入っているが、「低下し」や「先送りされる」などの節として取得しているため、適切に判定ができていない。これについては、単語の原形による部分一致などにより改善できないかを検討していきたい。

4.3.3 事象の同一性判定

表 9 に LDA により推定された各事象に対するトピックの確率を示す。今回の予備実験においてトピック数は 9 であるため、すべてのトピックを等しく含む事象の場合、トピックの確率 $P(t_i|I)$ は 0.11 となる。表 9 を見ると、どの事象も 0.11 前後の確率であり事象に対するトピックの割り当てがうまく行っていないことが伺える。これについては各事象における単語数が数単語と少ないことが原因と考えられる。

構築されたトピックを事象の同一性判定及び統合を行ったと

(注5) : Cabocha <https://code.google.com/p/cabocha/>

表 5 データセットとして利用した商品 (2013 年 10 月-2013 年 12 月)

商品名	取り扱い会社
アジア REIT・リサーチ・オープン (毎月決算型)	三井住友トラスト・アセットマネジメント
グローバル・ソブリン・オープン (毎月決算型)	国際投信投資顧問
J-REIT リサーチオープン (毎月決算型)	三井住友トラスト・アセットマネジメント
日本株式インデックス・オープン	三井住友トラスト・アセットマネジメント
SMT AMダウ・ジョーンズ インデックスファンド	三井住友トラスト・アセットマネジメント

表 7 LDA により構築されたトピック

ID	上位 10 単語
0	株式市場, 量的金融緩和, 証券優遇税制廃止, 公募増資, 貿易面, 景気回復, 先進主要国主導, インフラ整備, 下, トラック
1	米国, 長期金利, 緩やか, 最近, ユーロ高, 政府機関, 米欧主要企業, 景気回復, 一方, 中国
2	魅力, 一時, 足元, 通り, ブラジル, 多く, 今後, 衣食住, 年内, 先送り観測
3	公募増資, 新規上場, 上値, 一部, 今回, 貿易赤字, ユーロ圏, 製品, 無い, 運用方針
4	根強い, 欧米株式市場, 資金調達, 米国, 下記, 新興国, 建機, 資産買取規模, 現行, 規模
5	高水準, 配当利回り, その後, 需給悪化, 足元, 米国議会, 法案, 日経平均株価, 一部政府機関, 海外移転
6	投資対象, 景気, 国内, 投資口価格上昇, シナリオ, 投資削減, 注目点, 行方, 米議会, 国内需要
7	道路, 財政協議, 当面, 同問題, 米財政協議, 債務上限, サンテック REIT, 国内, 金利, 後半
8	量的金融緩和, 下値, 一巡, 悪材料, 情報提供, 円安効果, 大型商業 REIT, 今回, 当ファンド, 財政協議

表 8 抽出された因果関係の例

0. ((量的金融緩和, 米国, 先送り観測, 同問題), (合意, 縮小, 好感され), -) → ((投資信託), (経由し), +)
しかし、半ば以降は、同問題の合意や、米国の量的金融緩和縮小の先送り観測などが好感されたことを受けて、投資信託を経由した資金流入などから持ち直す展開となり、結局、前月末比小幅下落の水準で月内の取引を終えました。
1. ((米国, 量的金融緩和, 開始時期), (縮小, 早まる), -) → ((長期金利), (上昇, 転じ), +)
香港市場およびシンガポール市場は米国で量的金融緩和の縮小の開始時期が早まるとの観測から、長期金利が上昇に転じたことを受けて軟調に推移しました。
2. ((賃金カット, 負債圧縮, 緊縮財政, バブル崩壊後), (), +) → ((国内需要, 当時, 日本, 大幅), (消費, 投資, 減少し), +)
当時の日本は、バブル崩壊後の負債圧縮や緊縮財政、賃金カットなどを受けて、消費や投資などの国内需要が大幅に減少しました。
3. ((米国, 連邦政府機関, 一部), (縮小, 先送りされる), -) → ((長期金利), (低下し), +)
前半は、米国の連邦政府機関の一部の縮小が先送りされるとの観測から長期金利が低下したため、市場は緩やかに上昇する展開となりました。
4. ((投資削減, 鉱山企業), (), +) → ((景気, 一方, 中国, ブラジル, 下), (支えている, 鈍化してい), +)
一方で中国やブラジルなどの新興国では病院建設や道路などのインフラ整備が景気を下支えているものの鉱山企業の投資削減を受けて建機やトラックなどの需要が鈍化しています。
5. ((根強い), (), +) → ((下値, 高水準, 魅力, 配当利回り, 投資対象), (享受できる, 限定され), +)
しかし、政府・日銀による政策効果の浸透が引き続き期待されることや、また、不動産価格の上昇が期待される中、中長期的な観点では、高水準の配当利回りを享受できる投資対象としての魅力は根強いことから下値も限定され、底堅い展開を見込んでいます。

表 9 事象 0 から事象 9 に対するトピックの確率

ID	トピック 0	トピック 1	トピック 2	トピック 3	トピック 4	トピック 5	トピック 6	トピック 7	トピック 8
0	0.107	0.107	0.126	0.107	0.107	0.107	0.126	0.107	0.107
1	0.143	0.105	0.124	0.105	0.105	0.105	0.105	0.105	0.105
2	0.109	0.109	0.109	0.109	0.129	0.109	0.109	0.109	0.109
3	0.107	0.126	0.107	0.107	0.107	0.107	0.107	0.107	0.126
4	0.103	0.121	0.103	0.103	0.103	0.103	0.140	0.121	0.103
5	0.103	0.103	0.103	0.121	0.121	0.103	0.140	0.103	0.103
6	0.109	0.129	0.109	0.109	0.109	0.109	0.109	0.109	0.109
7	0.107	0.107	0.126	0.107	0.107	0.107	0.126	0.107	0.107
8	0.105	0.105	0.124	0.124	0.105	0.124	0.105	0.105	0.105
9	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.129

ころ、同一性判定・統合前に 67 あった事象は同一性判定・統合後は 27 となった。以下に統合が行われた事例を示す。

A:({ 投資口価格上昇, 足元 }, {}, {}, +)

事象 A のトピックの確率: [0.0000, 0.0000, 0.0150, 0.0000,

0.0000, 0.0000, 0.0150, 0.0000, 0.0000]

事象 A が抽出された文: 足元の投資口価格上昇を受けて、新規上場、公募増資による資金調達が発達しており、着実に市場規模は拡大しています。

B:({ 投資口価格上昇, 足元 }, {}, {}, +)

事象 B のトピックの確率: [0.0000, 0.0000, 0.0150, 0.0000, 0.0000, 0.0000, 0.0150, 0.0000, 0.0000]

事象 B が抽出された文: 足元の投資口価格上昇を受けて、新規上場、公募増資による資金調達が活発化しており、J-REIT の市場規模は着実に拡大しています。

この事例はコサイン類似度 1.0 により統合が行われた。この 2 つの事象については事象が抽出された文が非常に類似している。実際に月次レポートを確認していると、同一商品について異なる月で非常に類似した記述が行われていることが時折見受けられ、これは月次レポートにおける記述の特徴であると考えている。

また、別の事例を示す。

A: ({ 現行, 規模, 当面, 資産買取 }, { 継続される }, {}, +)

事象 A のトピックの確率: [0.0103, 0.0000, 0.0000, 0.0000, 0.0288, 0.0000, 0.0000, 0.0103, 0.0000]

事象 A が抽出された文: またバーナンキ米連邦準備理事会 (FRB) 議長は年内の資産買取規模の縮小に言及していたものの財政協議の行方を見極めることを優先するために当面現行の規模での資産買取が継続されるとみられることから長期金利は一定のレンジ内で推移するとみています。

B: ({ 米国, マクロ景気指標, 堅調 }, {}, {}, +)

事象 B のトピックの確率: [0.0126, 0.0000, 0.0000, 0.0000, 0.0314, 0.0000, 0.0000, 0.0000, 0.0000]

事象 B が抽出された文: 香港市場およびシンガポール市場は、米国の堅調なマクロ景気指標などを受け量的金融緩和の縮小懸念が強まったことから長期金利が上昇、月前半は軟調に移りました。

この事例においてはコサイン類似度 0.9472 により統合が行われた。この 2 つの事象については、トピック 0 及びトピック 4 の確率が高いため、コサイン類似度が高くなった。しかし、これらの記述を見ると、記述している内容は異なり本来であれば統合されるべきでない事象である。表 7 よりトピック 4 は米国、現行、規模などの単語を高い確率に持つトピックである。事象 A が「現行」、「規模」を事象 B が「米国」をそれぞれ含んでいたため、トピック 4 の確率が高くなりこの 2 つの事象が統合されたと見られる。

これはトピックの構築がうまく行えていないこと、また、事象に含まれる単語数が少ないため「米国」など一つの単語が事象におけるトピックの確率を求める上で大きな影響を与えてしまうことが原因であると考えられる。これについては、記述の周辺の語を考慮したり、また、経済用語辞書などを用いて事象の語を拡張するなどを今後検討していきたい。

最後に、トピックによる統合がうまく行えた事例を示す。

A: ({ 量的金融緩和, 米国, 開始時期 }, { 縮小, 早まる }, {}, +)

事象 A のトピックの確率: [0.0314, 0.0126, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000]

事象 A が抽出された文: 香港市場およびシンガポール市場は米国で量的金融緩和の縮小の開始時期が早まるとの観測から、長期金利が上昇に転じたことを受けて軟調に移りました。

B: ({ 量的金融緩和, 後, 若干, FOMC(米連邦公開市場委員会), { 開催され, 縮小, 決定され, 反発し }, {}, -)

事象 B のトピックの確率: [0.0288, 0.0103, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0103]

事象 B が抽出された文: 17、18 日に開催された FOMC (米連邦公開市場委員会) において量的金融緩和の縮小が決定された後は悪材料出尽くしから若干反発したものの、その後も長期金利の上昇が続いたことから反発も限定的なものにとどまり、月前半の下落分を取り戻すまでには至りませんでした。)

この事例においてはコサイン類似度 0.9472 により統合が行われた。それぞれトピック 0 を高い確率で持っており、表 7 よりトピック 0 は量的金融緩和を高い確率に持つトピックであるためである。トピックを利用して適切に事象の同一性判定が行えたと言える。

4.3.4 出力された因果関係図

図 4 に事象の同一性判定・統合を行い最終的に獲得した因果関係図を示す。プログラムより gexf ファイルとして出力し、グラフ可視化ソフト Gephi^(注6)を用いて可視化を行った。

図 5 にアジア REIT リサーチオープン周辺の因果関係を拡大したものを示す。因果関係図より量的金融緩和や実質のゼロ金利政策が国内の資金調達に影響を与え、アジア REIT リサーチオープンに影響を与えた因果関係があったのではないかと伺うことができる。しかし一方で、賃金カット、貿易赤字など資金調達に影響を与えなさそうな単語などが事象に含まれている。これらは過剰に事象を統合してしまったためであると考えられ、今後、事象の統合の手法についてさらなる検討が必要である。

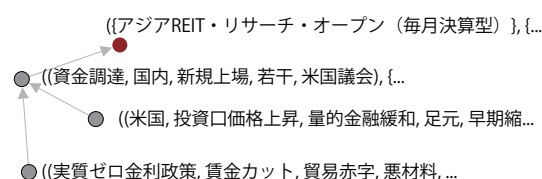


図 5 アジア REIT リサーチオープン周辺の拡大

5. ま と め

投資信託の月次レポートから因果関係図を構築し、複数の投資信託の比較検討を支援する手法について検討し提案した。また、投資信託の特徴である基準価額を用いた因果関係の抽出、因果関係の事象の極性の考慮、また、確信度・支持度を考慮した因果関係の表現について提案を行った。予備実験では手がかり表現を用いた因果関係の抽出を行い、LDA で構築したトピックを特徴量とし事象の同一性判定・統合を行い、得られた因果関係図を示し考察を行った。

(注6) : <https://gephi.org/>



図 4 出力された因果関係図

6. 今後の課題

6.1 事象のモデルの構築

予備実験の考察 (§4.3.3) より事象のトピックの確率が各事象の特徴を示すことができていない課題が明らかになった。事象に含まれる単語数が数単語と少ないことから、トピックの構築がうまく行えず、また 1 単語の有無がトピックの確率の推定に及ぼす影響が大きいためトピックの確率が適切に推定されなかったと考えられる。これに対して、経済用語辞書などの記述を利用して事象の記述を拡張する、記述の周辺単語を考慮するなどして、事象の単語数を増やすことで、LDA が事象の特徴を表すトピックを構築することができないか今後検討していきたい。また、Mikolov ら [6] はコーパスを用いて周辺単語を元に各単語の意味を表すベクトルを推定する手法を提案している。生成されたベクトルは単語の意味についての加減算が行える特徴を持つ。これらの LDA 以外の手法についても、事象の特徴を表すモデルを構築する上で利用できないか併せて検討していきたい。

また、例えば J-REIT の商品内でなど同一カテゴリ内で個別の商品を比較したいニーズがある。現在の事象のモデルの元で、同一カテゴリ内での商品の差を表現できるかを予備実験により評価した上で現在のモデルでは表現できない場合、それらの特徴を表現するモデル・手法を今後検討をしていきたい。

6.2 重要度を考慮した因果関係図の出力

予備実験で得た因果関係図 (図 4) では抽出した因果関係を全て出力したため、煩雑にエッジが張られ見づらく、理解支援の観点からは課題がある。予備実験では、確信度・支持度を考慮できていなかったが、今後、確信度・支持度を考慮し追加実

験を行い、理解支援に適切なグラフを出力できるかどうかを実験・評価していきたい。

また、現在、因果関係の重要度は出現頻度とその要因が記述されたレポートの月における基準価額の変動の大きさを指標として求めることを検討しているが、出現頻度は極めて低いが、その商品を理解するのに重要度が高い因果関係がないかなど、それらの指標で適切であるかを今後検討していきたい。

6.3 手法・システムの評価

システムが出力したモデルが実際の経済事象と商品の基準価額の変動を適切に表現できているか及び、システムが出力した因果関係図を利用したことにより投資商品の選択の理解が促されたかについての評価を今後行なっていきたい。

謝 辞

本研究の一部は、科研費（課題番号 25700033）と SCAT 研究費助成による。

文 献

- [1] 長尾 真. (2001). 「わかる」とは何か. 岩波新書
- [2] 坂地泰紀, 竹内康介, 関根聡, & 増山繁. (2008). 構文パターンを用いた因果関係の抽出. 言語処理学会第 14 回年次大会発表論文集, 1144-1147.
- [3] Ishii, Hiroshi, Qiang Ma, and Masatoshi Yoshikawa. "Incremental Construction of Causal Network from News Articles." *Journal of Information Processing* 20.1 (2012): 207-215.
- [4] Kawai, Hideki. "A Study on Search Methods for Knowledge Extraction and Structuring from the Web"
- [5] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.
- [6] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in Neural Information Processing Systems*. 2013.