

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments



馬 強 (Qiang MA, Ph. D.)

京都大学大学院 情報学研究科 准教授

(Associate Professor, Graduate School of Informatics, Kyoto University)

ACM IEEE 情報処理学会 電子通信情報学会 日本データベース学会

受賞：日本経営学会 中国・四国支部長賞 (1998年)

研究専門分野：観光情報学、投資情報学、情報栄養学 (情報補完) を中心にマルチメディア情報システム、Web マイニングの研究

あらまし NISA (少額投資非課税制度) などの政策支援もあって、「貯蓄から投資へ」のスローガンのもとで多くの人々が投資を行うようになってきている。しかしながら、商品に対する知識や理解に不安を覚え、投資に踏み切ることができない人がまだ多い。「関係」の整理と理解は人間知の活動において極めて重要であることに着目して、筆者らの研究グループでは、投資商品に対する理解や投資活動における意思決定を支援するため、インターネット上に公開されている情報を対象に、企業や事象の関係を分析する技術について研究開発を行っている。本稿では、筆者らが開発している、1) 事象間の因果関係を分析して投資商品の理解を支援する技術、および、2) 企業間の関係を明らかにする事象と株価の統合分析技術について紹介する。

1. はじめに

日本が超高齢化社会に突入しつつある。年金不安を緩和して安定な老後生活を送るためには、投資の重要性と必要性がますます高まってきている。「貯蓄から投資へ」というスローガンのもと、政府の政策支援もあって、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えている。しかしながら、一般社

団法人投資信託協会の「投資信託に関するアンケート調査報告書¹でも報告されているように、投資商品に対する理解に不安を覚え、投資に踏み切ることができない人がまだ多い。投資商品に対する理解支援が非常に重要な課題となっている。

Bush 博士が Memex 構想[1]で指摘されているように、「関係」の整理と理解は人間知の活動において極めて重要である。例えば、企業間の関係分析は、マーケティングや意思決定において重要な役割を持ち、組織間関係論の長年の研究対象となっている[2]。また、事象間の関係、とりわけ因果関係の分析は、過去の事象の理解支援と将来の予測や意思決定には必要不可欠である[3][4]。

そこで、筆者らの研究グループでは、関係マイニングのアプローチから投資活動の支援を試みている。本稿では、以下の二つの成果について紹介する。

- 事象間の因果関係分析技術：投信など投資関連商品の基準価額に影響を及ぼす要因を定量に分析し、商品の比較検討を支援する技術について研究開発を行っている[4]-[7]。
- 企業間の関係分析技術：事象と株価を統合分析して企業間の関係を推定し、公式サイトなどのテキスト情報に記述されていない暗黙な関係も発見可能とするとともに、事象の株価への影響をより正確に予測可能となる技術について研究開発を行っている[8][9][10]。

¹ <https://www.toushin.or.jp/statistics/report/research2014/>

2. 関連研究

2.1 因果関係分析

テキストから因果関係を抽出する研究が多くある。例えば、酒井らが業績に影響を及ぼす要因を決算短文から抽出する手法を提案している。彼らはポジティブ・ネガティブ表現を「により」や「から」といった手がかり表現を用いて収集することで、要因抽出の精度を向上させている[11]。また、坂路らは機械学習手法を用いて原因と結果を含む文を抽出する手法を提案している[12]。

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments

これらの既存研究は、金融関連テキストにおいてテキストデータの因果関係を適切に把握する際に極めて有用であるが、因果関係の強さや要因の影響範囲などを定量的に分析する研究がまだ少ない。

2. 2 企業間関係分析

テキストデータから企業間ネットワークを抽出・分析し、知見を得る研究は盛んに行われている。例えば、金らは Web 上に存在している情報から企業間の関係を明らかにし、企業ネットワークを構築する手法を提案している[13]。金らの研究は企業間の関係性を導き出す情報として、Web 上のテキスト情報のみを対象としており、抽出する関係性の対象も提携関係と訴訟関係のみに絞っているが、本研究では関係性の対象を取らず、企業の業績に焦点を当てて関係性があるかどうかを判断する点が異なる。

Michael ら[14]や Aliakbar ら[15]は、サッカーのクラブチームの試合結果がスポンサー企業の株価に影響を与えることについて分析を行っている。これらの研究では、企業間の暗黙的な関係の重要性を指摘してきたが、暗黙に關係する企業を、スポンサー関係をもとに手動で与えている。本研究では、ニュースと株価のデータをもとに企業間の関係を自動発見する点が異なる。

3. 事象の因果関係分析による投資支援

投資信託商品は、専門知識がなくても投資が出来ることから、金融商品として大変人気がある。しかし、投資信託に関するアンケート調査報告書(1節を参照)によると、投資信託の現在保有率は16%に留まり、保有層も60代以上が約半数で、高齢層が中心である。また、非保有者の多くは投資信託商品の理解に不安を感じ、投資に踏み切ることができないことも報告されている。投資信託商品の特徴を理解するための重要な情報源である月次報告書を、48%の投資家がまったく読んでおらず、きちんと読み込んだ投資家も全体の8%ほどに過ぎない現状から、適切な情報利用もできていないと言える。

筆者らは、利用者がどういった情報を理解したいのかを探るため、クラウドソーシングを用い、投資経験

が異なる183名ワーカーに対し調査を行った[5]。調査では、彼らが金融緩和や経済政策のようなマクロ要因だけではなく、個別企業の業績などのミクロ要因も理解したいことが分かった。したがって、マクロ的要因とミクロ的要因の両方を分析することが理解支援に必要である。しかしながら、市場の変動要因の理解支援に関する研究の多くは、マーケット全体に影響を及ぼすマクロ的な要因のみにフォーカスしている。

要因の基準価額への影響の度合いは、投資信託商品ごとに異なったり、投資信託商品の構成銘柄の入れ替えによる変化したりするため、諸々の要因の影響を時系列に定量的に観察する必要がある。しかしながら、既存のテキストマイニングによる要因分析の手法の多くは、その影響を定量的に分析していない。特に、影響の時系列変化を考慮した要因分析の研究が少ない。

そのため、筆者らは、状態空間モデルを適用して商品理解を深めるための要因の定量分析手法を開発している。状態空間モデルは、状態方程式と観測方程式から構成される汎用的な時系列解析の手法であり、高い柔軟性を持つことから、金融やマーケティングなど様々な分野で利用されている。

図1は、開発手法の概要を示す。テキストデータとして、投資信託商品の運用報告書である月次報告書と日経QUICKニュース社から提供されたニュース記事を用いる。まず、これらのテキストデータから要因を抽出し、日経シソーラスを用いて要因を分類する。開発手法では、日次データと関連づけることが可能である「マクロ変数要因」、経済指標や国内外の政策のようにある時点から影響を与える「マクロ干渉要因」、業績などの個別企業に関わる「ミクロ要因」に分類する。次に、要因の発生日と基準価額のトレンド変化を検知して、これらの要因の影響範囲を推定する。さらに、要因の多重共線性も考慮して必要な要因を説明変数として選択し、状態空間モデルを構築する[5]。

上述したように、要因の定量分析は基本的に状態空間モデルを用いて行うが、状態空間モデルを構築する際、要因の選別および要因の有効時間(影響範囲)の推定も行う点が独創である。

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments

- 変数選択では、要因の種類および要因間の多重共線性を考慮して、モデル構築に適切・必要な要因を選択する手法を実現している。
- 要因の時系列変化を考慮した状態空間モデルを構築するために、動的計画法を用いた基準価格のトレンドを分割する。状態空間モデルをトレンド局面ごとに構築することで、その局面で影響を与えている要因のみで分析を行い、要因の影響度をより正確に推定できるとともに、要因の影響の時系列変化を明らかにすることが可能となる。

評価実験では、2016年1月から10月の期間の国内株式を対象とした投資信託商品を用いた[5]。実験結果から、要因の抽出および分類を高い精度で行えることを確認できた。また、トレンドごとに状態空間モデルを構築した状態空間モデルの誤差が、トレンドを考慮しない場合より小さいことを確認できた[5]。

さらに、図2のような投資商品を比較するインターフェースも開発している[4][7]。図2では、US、アジア、そして日本のREIT商品を比較している。東京オリンピックが日本REITの商品に正の影響を与えているが、アジアREITのもっとも重要な要因はシンガポールGDPであることが分かる。開発したユーザインターフェースを利用した場合、既存サービスの半分の時間で商品の特徴を理解可能となったことをユーザ実験の結果から分かった[7]。

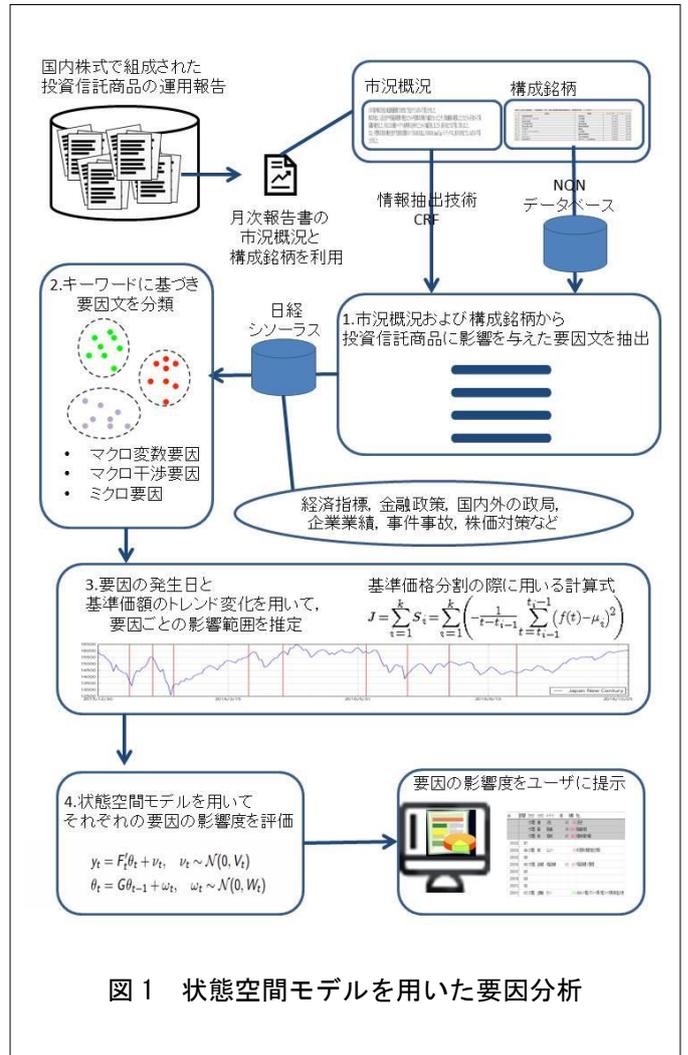


図1 状態空間モデルを用いた要因分析

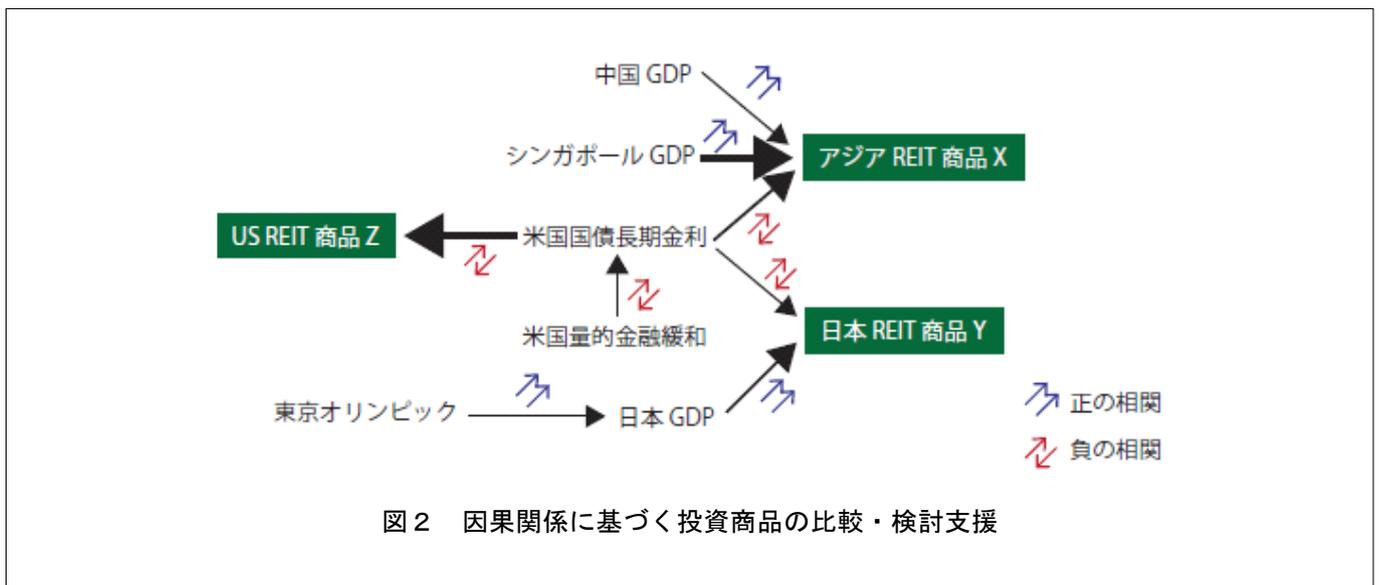


図2 因果関係に基づく投資商品の比較・検討支援

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments

4. 企業間の関係分析

本研究では、企業同士の関係を分析することで、企業の競争力の分析や戦略決定、個人投資家の企業の成長性の分析や投資企業選定の支援を行う。具体的に、企業間の関係をニュースイベントと連動した株価の相関から推定する手法[8][10]と、ヘテロトピックモデルを用いて確率的学習する手法[9]について研究開発している。

企業間には様々な関係が存在する。グループ会社、子会社といった Web ページや広報資料などで明示されている関係もあれば、スポンサー関係や取引先など明らかにされていない暗黙的な関係もある。企業の Web サイトやニュース記事などを調べて前者の関係を分析できるが、後者の暗黙的な関係の発見が難しい。企業間の関係性を分析する研究は既に行われているが、インターネットで得られるテキスト情報から明示的な関係のみの分析を行っているものが多い[2][13]。これに対して、本研究の提案手法は、公式サイトなどのテキスト情報に記述されていない暗黙的な関係も発見可能であるとともに、事象の株価への影響も予測できる。

4.1 株価動向を用いた企業間関係分析

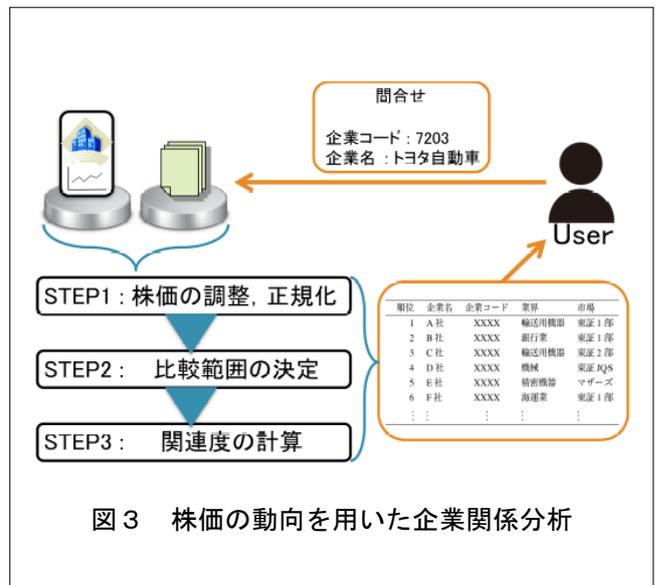
株価が企業の業績を反映することが多い。株価が上昇しているときは企業の業績もよく、株価が下落しているときは企業の業績も悪くなることが多い。また、業績に影響を及ぼす関係が存在すれば、両企業の株価に相関の高い動きが現れることが多いと考えられる。例えば、ある企業が自動車の開発を行った際に、別の企業の部品を使用していれば、その部品の売れ行きは自動車の売れ行きに左右される。開発した自動車が売れて、メーカーの業績が良くなり株価が上昇すれば、部品を供給している企業の業績も良くなり、株価が上昇する。したがって、株価の動きの関連性を調査することで、企業間の関係性を明らかにすることが可能である。

しかしながら、企業の株価は、その企業の業績のみではなく、業種や市場全体の影響も受ける。そこで、筆者らは、企業自身の要因に影響される株価を実価、業種の影響度合いを業種指数、市場の影響度合いを市場指数として、企業の株価はこの三つの要素からなる株価の合成モデルを構築する。株価の合成モデルを用

いて求められた実価の相関を分析して、企業間の関係を発見する(図3)。

実価を用いて企業間の関係を分析する際、まず、企業の実価の前日比を求めて正規化し、異なる企業の株価の値幅の差をなくす。次に、企業間の関係が動的に変化することを考慮して、企業に關係するニュースイベントをトリガーとして株価の比較する範囲を決定する。これによって、企業間の関係の変化を捉えるとともに、計算量の削減効果もある。そして、比較範囲の株価データ(系列)を入力とし、SAX法(Symbolic Aggregate approximation) [19]を用いて企業間の関連度を推定する。

評価実験[10][11]では、合成モデルを用いて算出された実価を用いた手法が企業間の関係をより正確にとらえることが分かった。また、ニュースイベントを用いた方が、計算量を削減できるほか、関係の検出精度の向上にも効果があったことが確認できた。



4.2 ヘテロトピックモデルを用いた企業関係学習

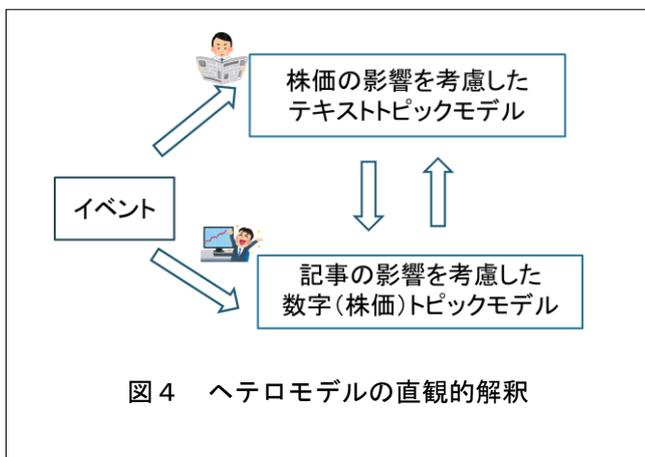
数値データである株価とテキストデータであるニュース記事を用いて、株価予測と企業関係発見を統合的行えるヘテロトピックモデルを開発している[9]。このモデルを用いると、事象に応じた投資先企業の選定や株価予測、および企業間ネットワーク分析などの支援を行える効果が期待できる。

ヘテロトピックモデルは、ニュース記事の単語はト

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments

ピックが決定されて生成されるトピックモデルである LDA (Latent Dirichlet Allocation) [16] を拡張した確率生成モデルである。図 4 に示しているように、筆者らは、事象 (イベント) はニュース記事 (その内容) と株価 (上昇・下落) の両方に影響を及ぼすことに着目して、ニュース記事のテキストベースのトピックモデルと株価の数字ベースのトピックモデルを統合して扱うヘテロトピックモデルを構築している。



LDA をはじめとするトピックモデルでは、観測データとして用いられるのはテキストデータのみであり、テキスト以外の観点を含めた多角的な分析を行えない欠点がある。ニュースは株価に影響を与える場合が多いことから、株価への影響を考慮して経済ニュースのトピックを、より正確に推定できる効果が期待できる。同様に、事象の株価への影響を推定する際、事象を報道するニュース記事の内容を考慮することで、推定精度の向上につながる効果を期待できる。

ヘテロトピックモデルの特徴を以下に列挙する。

- トピックが企業のグループにどのくらい影響を与えているかを示す潜在変数を導入している。この潜在変数を学習・推論することで、事象による株価への影響の強さだけでなく、企業間の関係を推定可能となる。
- スイッチ変数を導入し、トピックの株価への影響の有無をモデリングしている。ニュース記事内の単語のうち、企業の株価に影響を与えるものと与えないもので考慮する確率分布を切り替えることができ、テキストと数字データを統合的に分析できる。

ヘテロトピックモデルのパラメータは、ギブスサンプリング法を用いて学習する。筆者らは、ある経済事象が発生したときにニュース記事の生成過程、それによって発生する株価の変動過程を明らかにすることを目的としている。よって、まず、ニュース記事において単語をトピックに割り当て、トピック分布のパラメータを更新する。次に、更新されたパラメータを用いて、ニュース記事が株価に与える影響を考慮し、株価の変動が類似しているグループの分布 (トピック分布) のパラメータを更新する。これを繰り返すことにより、ニュース記事が企業の株価に与える影響を考慮したヘテロトピックモデルのパラメータを学習する。

予備実験ではスイッチ変数の効果を確認できたが、データ数が少ないためトピックの推定では特定のトピックに偏った結果となった。また、学習の効率の向上が課題であることも分かった。今後、これらの課題について取り込んでいく予定である。

5. まとめ

関係の整理と理解は、知の活動においてきわめて重要である。そのため、筆者らは、インターネット上に射影されている情報を対象に、事象やエンティティ間の関係を分析して社会活動に利活用する技術について研究開発を行っている。特に、投資活動を支援するための事象間の因果関係や、企業間関係の分析について研究を推進している。

超低金利時代に老後の年金生活の不安を解消するため、投資が必要不可欠な手段とされている。しかしながら、投資に関する知識が乏しく、投資商品を理解できないといった理由から投資を敬遠する人がまだ多い。情報技術を活用していかに投資活動を支援するかが大きな課題となっている。本稿では、ユーザが自分で投資商品を比較分析して購入することを支援するための研究について紹介しているが、投資の達人の行動を複製して投資を行うソーシャルトレーディングサービスにおける複製するトレーダの選別の支援に関する研究も遂行している [17][18]。ソーシャルトレーディングサービスは、ユーザが自分でテクニカル分析やファンダメンタル分析を行って投資商品を選別する必要がないため、注目を集めている新しい投資手法である。

投資支援のための関係マイニング技術の開発

A Study on Relationship Mining for Supporting Investments

今後、これらの成果を踏まえて、関係マイニングの情報基盤を確立するとともに、投資支援のシステムの実用化へ向けて研究を進める予定である。

参考文献

- [1] Bush, V.: As We May Think. Atlantic Monthly, Vol. 176, No.1, pages 101-108, 1945.
- [2] 山倉健嗣: 組織間関係と組織間関係論, 横浜経営研究, Vol. 16, No.2, pp.166-178, 1995.
- [3] Ishiii, H., Ma, Q., and Yoshikawa, M.: Incremental Construction of Causal Network from News Articles, Journal of Information Processing, Vol.20, No. 1, pp.207-215,2012.
- [4] Awano. Y., Ma, Q., Yoshikawa, M.: Causal Analysis for Supporting Users' Understanding of Investment Trusts, iiWAS 2014, pp. 524-528, 2014.
- [5] 大西恒彰, 馬強: 状態空間モデルの拡張による投資信託商品の要因分析, DEIM2017,2017.
- [6] 大西恒彰, 馬強: 月次報告書と基準価格を用いた投資信託商品の要因分析, DEIM2016, 2016.
- [7] 粟納裕貴, 馬強, 吉川正俊: 因果関係分析に基づく投資関連商品の比較手法の検討, DEIM2014, 2014.
- [8] Baba, S. and Ma, Q.: Analyzing Relationships of Listed Companies with Stock Prices and News Articles. DEXA2016, pp.27-34, 2016
- [9] 馬場慧, 馬強: 株価とニュースの統合分析のためのヘテロトピックモデル DEIM2017
- [10] 馬場慧, 馬強: 株価とニュース報道を用いた上場企業の暗黙関係の発見, DEIM2016, 2016
- [11] Sakai, H. and Masuyama, S., Assigning Polarity to Causal Information in Financial Articles on Business Performance of Companies, IEICE Transactions on Information and Systems, vol.E92-D, No. 12, pp. 2341-2350, 2009.
- [12] Sakaji, H., Sekine, S, and Masuyama, S.: Extracting Causal Knowledge Using Clue Phrases and Syntactic Patterns, Proc. of 7th International Conference on Practical Aspects of Knowledge Management, pp. 111-122, 2008
- [13] 金英子, 松尾豊, 石塚満: Web 上の情報を用いた企業間関係の抽出, 人工知能学会論文誌, Vol. 22, pp. 48-57, 2007.
- [14] Hanke, M. and Kirchler, M.: Football championships and jersey sponsors' stock prices: an empirical investigation, The European Journal of Finance, Vol. 19, pp. 228-241, 2012.
- [15] Ramezani, A., Mardani, H., Emamgholipour, M. and Mar-dani, S.: The Effect of the Results of Football Champions League Games on Sponsors' Stock Prices: Evidence from Iran, World Applied Sciences Journal, Vol. 20, pp. 102-106, 2012
- [16] Blei, D.M., Ng, A.Y., and Jorda, M.I.: Latent dirichlet allocation, Journal of machine learning research, Vol. 3, No. Jan, pp.993-1022, 2003.
- [17] Lee, W., Ma, Q.: Whom to Follow on social trading services? A system to support discovering expert traders. ICDIM 2015, pp.188-193, 2015.
- [18] 竹田創, 馬強: ソーシャルトレーディングサービスにおけるトレーダの特徴分析, DEIM2017, 2017.
- [19] Lin, J., Keogh, E., Lonardi, S. and Chiu, B.: A Symbolic Representation of Time Series with Implications for Streaming Algorithms, DMKD03, pp.2-11, 2003.

この研究は、平成24年度SCAT研究助成の対象として採用され、平成25～27年度に実施されたものです。