

令和 5 年 5 月 9 日現在

機関番号：12601

研究種目：若手研究

研究期間：2020～2022

課題番号：20K20130

研究課題名（和文）異質情報ネットワークマイニングによる金融市場における情報アクセスの向上

研究課題名（英文）Improving Information Access in Financial Markets through Heterogeneous Information Network Mining

研究代表者

久野 遼平（Hisano, Ryohei）

東京大学・大学院情報理工学系研究科・講師

研究者番号：60725018

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究はビッグデータをノレッジグラフとして統合し、分析技術を開発することで、市場参加者が広範囲に必要な知見を検索するための基盤技術を創出することである。初年度では、大規模なニュースデータを分析する手法を開発すると同時に、ヘルムホルツホッジ分解を用いて外交政策の影響ネットワークの階層構造を抽出することに成功した。二年目では、各国のスマート制裁リストにエンティティを追加するタイミングのデータを分析して、外交政策のパワーバランスを光に当てた。最終年度では、法律・判例を対象にした法の構造分析をさらに発展させ、銀行送金データを分析する新たな角度から研究を進めた。成果は国際ジャーナルに複数論文を掲載した。

研究成果の学術的意義や社会的意義

金融市場はもちろんのこと、情報が意思決定の重要な基盤となる領域において、意思決定者が社会の持続的発展に貢献するためには、参加者が広範囲にわたって十分に正確な情報に基づいて意思決定を行うことが必要不可欠である。そこで本研究では、ビッグデータの集合体を異種情報ネットワークとして統合し、ネットワークやテキストマイニングを基盤とした分析技術を開発することで、市場参加者が多様な情報の中から必要な知見を合理的かつ広範囲に検索するための基盤技術を創出する。本研究によって、参加者が無知によって望まない判断をすることを防ぐことができるようになる。

研究成果の概要（英文）：This study aims to integrate big data as a knowledge graph, develop analysis techniques, and create a foundational technology for market participants to search for necessary insights widely. A method for analyzing large-scale news data was developed in the first year. The hierarchical structure of the influence network of foreign policy was successfully extracted using the Helmholtz-Hodge decomposition. In the second year, analyzing data on the timing of adding entities to each country's smart sanctions list highlighted the power balance of foreign policy. In the third year, the structural analysis of law and precedent was further developed, and research was carried out from a new perspective of analyzing bank transfer data. As a result, a paper was published in several international journals.

研究分野：データサイエンス

キーワード：ネットワーク学習

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

金融市場が社会の持続的発展に貢献するには、市場参加者が広範囲にわたって十分に正確な情報に基づいて意思決定を行うことが必須である。そこで本研究ではビッグデータの集合体を異質情報ネットワークとして統合し、ネットワークやテキストマイニングをベースにした分析技術を開発することで、市場参加者が多角的な情報の中から必要な知見を、合理的かつ広範囲に検索するための基盤技術を創成する。本研究課題によって市場参加者が適切な情報に接しないことで望まない判断をすることを防ぐことが可能になる。

### 2. 研究の目的

本研究ではプロの投資家が用いる企業に関するビッグデータを異質情報ネットワークとして統合し、先進的なネットワークマイニングとテキストマイニング技術を開発することで、市場参加者が膨大な情報の中から必要な知見を、真偽性に惑わされずに、合理的かつ広範囲に検索するための基盤技術を創成する。異質情報ネットワークを用いるのは複数ソースからの多角的な情報（ネットワーク、テキスト、数値情報）を総合的に分析するためである。本研究の核心をなす学術的な問いは「複数ソースの情報を統合し、利活用した分析体系を創ることによって安全かつ範囲にわたった情報アクセスはどこまで可能になるか？」である。本研究によって、市場参加者が世界中の情報に広範囲にわたってアクセスし、的確にグローバルな投資判断をできる社会を目指す。ただし、本研究の手法は何も金融市場にのみ意味があるものではない。そのためデータの取得状況によって分析領域は柔軟に適応させていく。

### 3. 研究の方法

次の3つの研究課題に取り組むが発想自体は幅広い分野に適用できるため分野に関しては柔軟に対応する。1. ESG投資に影響を与える重要な負のニュース予想とソースの信頼性: 本研究者の過去研究では「重要なニュース」と「浅薄な嫌疑」を峻別していない。それは以前使用したデータセットはニューステキストデータそのものではなくイベントデータであったため、負のニュースではあることは間違いないが、程度は考慮しようがなかったためである。そのため本研究ではニューステキストデータ自体を扱った研究を実施する。またたとえプロの投資家が使用するデータであったとしても、ニュースソースの中には信頼性が低いものも含まれている。そこで本研究ではまず、(i) ニュース発信元・日付情報を考慮しつつ(ii) 単語レベルの分析を行うことによって「重要なニュース」(逮捕された、裁判になった)に対して「浅薄な嫌疑」(見込みである、憶測が出ている)を峻別しながら、(iii) テキストから企業同士の関係、根拠文、意見文を抽出し、総合的に分析することで、(a) どのニュースソースが重要なファクトを一早く報道したか、(b) ニュースソースの傾斜具合、(c) ニュースソースの信頼性スコアの計算を行う。こうしたテキストから分析した情報を元に改めて2.(1)の研究を見直し、重要な負のニュースを予想できるように発展させる。2. 異質情報ネットワークを用いた企業の財務状態の推定: 近年ネットワーク上のあるノードやサブネットワークの状態を周囲にある情報を集約化することで推定する Neighborhood Aggregation と呼ばれる方法が提案されている。本手法はあるノードを起点として周辺の情報を集約化する関数を構成することで特徴量を構成する手法である。近年研究のためネットワーク技術として改善できている点が多いのも関心をもっている理由であるが、本研究の文脈においてはある企業が財務上危機的な状況にある(倒産・大幅減益)際に異質情報ネットワーク上の周囲の情報に何等かのパターンがあるかどうか判定する問題に適用することができる。まずは本手法を異質情報ネットワークに拡張しノードの状態の予想精度の向上を図る。

複数ソースの情報の活用が重要なのは何も金融ドメイン(ESGリスク・スマート制裁予想)だけではない。例えば法分野においても法律・判例・事例など多角的な情報を加味して分析体系を構築する必要がある。また判決文は文章が構造化されているため通常のニュースデータよりははるかにデータを分析しやすいという利点もある。そこで法律・判例・事案などを異質情報ネットワーク上のノードとして定義し統合的に分析する手法についても研究を行う。

### 4. 研究成果

研究初年度は、大規模なニュースデータから、人々が物事を捉える方法の違いを分析する手法の開発に取り組んだ。ニュースデータはダウ・ジョーンズから購入し、データの整理作業にも従事した。手法については、荒っぽいプロトタイプは完成しており、学会誌にも投稿済みであるが、細かい点にまだまだ改善の余地があるため、その修正に取りかかっている。具体的には、テキストから情報を抽出する深層学習技術に関して、適切なドメイン適応手法を検討した。また、多角的なノレッジグラフを作成するために、今まで利用してこなかったようなデータも分析範囲に含めた。具体的には、各国のスマート制裁リストにエンティティを追加するタイミングのデータである。これらのデータを用いることで、どの国が先に動き、どの国が追従したかを抽出し、外交政策上のパワーバランスに光を当てることができるようになった。当該研究

では、ヘルムホルツホッジ分解を用いることで、外交政策の影響ネットワークの階層構造を抽出することに成功した。本研究は論文化し、国際発表も行った[1-2]。

研究の二年目は、前年度分析したスマート制裁リストを統合した。データの取得は価格の高騰もあり思い通りには進まなかったが、最終的には入手でき、計算環境の整備も終わったため、研究期間終了後も引き続き本データベースの構築に邁進していく予定である。技術面については、解釈性に焦点を当てた研究に従事した。これは、今までの研究では多角的なノレッジグラフを作成したものの、情報を完全に有効活用できていないため、モデルが何を示しているのかについて解釈性が弱いという課題があった。一方で、深層学習において解釈性は非常に注目されるトピックである。そのため、本研究でも、経済ノレッジグラフの中での解釈性に注目した。現在のモデルは試作段階にあるが、今後も引き続き開発に取り組む。最終的には COVID-19 の影響もあり、学会活動にはあまり参加できなかったが、本研究と関連する法の構造分析については、招待講演も受けた[3]。こちらにも、法律・判例・事例が含まれ、多角的な分析体系を構築するものであり、双方の研究のシナジーも期待できる。

最終年度である三年目は二年目に注目した法の構造分析をさらに発展させることにした。本課題名にある金融に直接関係するわけではないが、データの取得がそちらの方が進んだ点と本質的な問題意識は同じであるためそう判断した。具体的には法律であれば法律・判例・事案などが異質情報ネットワーク上のノードにあたりそれらを統合的に分析することが目標となる。対象データを変えたことによって法律やよりまとまった判決文など構造化されたテキストデータからの情報抽出など新たな角度からの研究にも取り組むことができた。これは金融ニューステキストなどにも応用ができるものであり、予期しない形で一歩大きく研究が進んだと言える。成果としては判決文の言語分析と法律・判例を対象にした法の構造分析に関する論文を投稿し国際ジャーナルに掲載された[4]。本論文に関してはオープンアクセスにしてある。法の構造分析で得た知見と昨年度取り組んだ解釈可能性に焦点を当てた深層学習のモデルと組み合わせることで金融データに関しても入手し新たな角度から分析をすることができた。研究成果については現在執筆中であるが、新たなデータを入手できたので、成果発表に関しては少し先送りする予定である。同時に金融に関しては銀行送金データという新たに入手できたデータの分析も行った。これに関しては論文成果を発表済みである[5]。

#### 参考文献

- [1] Ryohei Hisano, Hiroshi Iyetomi, Takayuki Mizuno, "Identifying the Hierarchical Influence Structure Behind Smart Sanctions Using Network Analysis", Lecture Notes in Computer Science 95-107 2020
- [2] Ryohei Hisano, Hiroshi Iyetomi, Takayuki Mizuno, "Identifying the Hierarchical Influence Structure Behind Smart Sanctions Using Network Analysis", Social Informatics, 12th International Conference SocInfo 2020
- [3] 久野遼平, テキストとネットワークマイニングによる法の構造分析, ネットワーク科学研究会 2021, 2021 年
- [4] Ryoma Kondo, Takahiro Yoshida, Ryohei Hisano, "Masked prediction and interdependence network of the law using data from large-scale Japanese court judgments", Artificial Intelligence and Law, 2022.10
- [5] 久野遼平, 長澤達也, 高橋秀, 近藤亮磨, 大西立顕, 銀行送金ネットワークの内在的構造と時間変化, 人工知能 38(2) 2023 年 3 月

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Kondo Ryoma, Yoshida Takahiro, Hisano Ryohei	4. 巻 None
2. 論文標題 Masked prediction and interdependence network of the law using data from large-scale Japanese court judgments	5. 発行年 2022年
3. 雑誌名 Artificial Intelligence and Law	6. 最初と最後の頁 None
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10506-022-09336-5	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 久野 遼平、長澤 達也、高橋 秀、近藤 亮磨、大西 立顕	4. 巻 38
2. 論文標題 銀行送金ネットワークの内在的構造と時間変化	5. 発行年 2023年
3. 雑誌名 人工知能	6. 最初と最後の頁 131 ~ 138
掲載論文のDOI（デジタルオブジェクト識別子） 10.11517/jjsai.38.2_131	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hisano Ryohei, Iyetomi Hiroshi, Mizuno Takayuki	4. 巻 1
2. 論文標題 Identifying the Hierarchical Influence Structure Behind Smart Sanctions Using Network Analysis	5. 発行年 2020年
3. 雑誌名 Lecture Notes in Computer Science, Social Informatics, 12th International Conference SocInfo 2020	6. 最初と最後の頁 95 ~ 107
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-60975-7_8	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 3件／うち国際学会 1件）

1. 発表者名 久野遼平
2. 発表標題 久野遼平、「テキストマイニングとは？政策の企画・立案における応用可能性」
3. 学会等名 経済産業省ワークショップ（招待講演）
4. 発表年 2022年

1. 発表者名 久野遼平
2. 発表標題 テキストとネットワークマイニングによる法の構造分析
3. 学会等名 ネットワーク科学研究会2021 (招待講演)
4. 発表年 2021年

1. 発表者名 Ryohei Hisano, Hiroshi Iyetomi, Takayuki Mizuno
2. 発表標題 Identifying the Hierarchical Influence Structure Behind Smart Sanctions Using Network Analysis
3. 学会等名 ecture Notes in Computer Science, Social Informatics, 12th International Conference SocInfo 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 久野遼平, 坂地泰紀
2. 発表標題 因果チェーンとネットワーク学習によるソースごとのCOVID19に関する捉え方の違いの抽出
3. 学会等名 東京大学金融教育研究センター・日本銀行調査統計局共催ビッグデータフォーラム (招待講演)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------