

研究報告書

「プロパティグラフによる社会データにおける常識の考慮」

研究期間：平成 29 年 10 月～平成 31 年 3 月

研究者番号：50151

研究者：久野遼平

1. 研究のねらい

本研究の狙いは企業、人、商品などに関する情報を複数ソースから集めて異質情報ネットワークとして分析することで、単一ソースの情報だけでは予想することや分析することが困難な諸課題に対して、予想精度の向上や未知の洞察の発見などの実証分析がどれくらい改善するかを検証することにある。無論これではあまりにも抽象的なため、本研究では特にテーマを ESG 投資(環境・社会・ガバナンス)に用いられる投資除外リストと制裁リストの予想に問題設定を絞った。この二つのテーマに絞った理由は両方とも国際的に関心が高まっている話題であること以外に個別にも背景がある。

ESG 投資は、(1)企業が自主的に公開した情報、(2)格付け会社の格付け、(3)過去のニュース情報などを用い行われてきた。(1)の自主公開した情報はエンロンの粉飾決算のように偽装可能性を排除することができない。(2)の格付け情報もサブプライム問題のように利益相反により必ずしも正しく行われているとは限らない。(3)の過去ニュースもメディアの調査報道によって未知の問題を炙り出すことができる反面、蓋然的に問題を抱えているが単に報道されていない企業を把握できないという問題を抱えている。つまり、ESG 投資を支える情報環境はまだ追いついていないのである。これに対し本研究では将来報道によって投資除外リストに企業が含まれる可能性の分析に焦点をあてた。これは従来研究がニュースが出てから反応するなど「泥棒を捕らえて縄を縛う」状況に限定されていたのに対して従来を越える革新的な一歩と呼べる。こうした ESG 投資にまつわる情報環境の問題を背景に本研究では ESG 投資を扱うことにした。

制裁リスト作成は(i)センシティブな特定ソース(米国財務省の対テロ戦争における国際銀行間通信協会データ)に頼った調査法がある一方で(ii)捜査員が複数の情報ソースからの情報を必要に応じて少しずつ手動で入手・検証する調査法もある。前者の方法に関しては捜査対象になりうる組織が特定のソースを回避することで足跡を残さないことができ、後者の方法は捜査員が手動で分析できるデータ量に限界があることに問題がある。本研究のように複数ソースのデータから自動で制裁対象を予想することができれば、捜査対象が足跡を残さないことを困難にすると同時に(ii)の地道な捜査に関して捜査員をサポートすることができるようになる。これが制裁リストも本研究で扱うことにした理由である。

2. 研究成果

(1) 概要

本研究で取り扱った問題設定を図1に示す。図1(a)にあるように本研究の目的は投資除外リストや制裁リストに新規に追加される企業を予想することにある。無論、リストをただ眺めているだけでは予想はできない。そこで本研究では図1(b)のように複数ソースからの情報を一つの異質情報ネットワークとして管理し、それを辿る方法を開発することで将来リストに追加される企業を予想するアプローチを採用した。図1(b)は Furukawa(2017)に記載がある北朝鮮制裁に関する国連捜査の過程を簡略化し図示したものである。特筆すべきは捜査員が複数の情報(登記簿、船舶、企業関係、所有関係)を手動で取得・検証し、積み重ねることで捜査を進めている点である。本研究の目指す所はこうした複数ソースを有効活用した捜査をデータの中から自動的にパターンを発見することでサポートすることにある。

図1(a)に対応するリストに関してはまず Dow & Jones が提供する ESG 投資に利用されるニュースデータを元に投資除外リストを構築した。その他にも各国や国際機関が発行する制裁リストのデータは公開情報で入手可能のため、そちらも収集し制裁リストのデータベースを構築した。図1(b)に対応するデータベースに関しては7種類の投資家が利用するデータと2種類のオープンデータを用い構築した(表1)。ネットワークとしてみた場合ノード数が5千万でエッジ数が4億のため豊富に情報が含まれていると言える。

図1(a)(b)のデータを用いラベル伝播法の亜種を用いることで ESG 投資で用いられる投資除外リスト予想に取り組んだ(Hisano et al. 2018)。結果をかいつまんで説明すると、複数ソースを用いる提案モデルは一般的に利用されるモデルに対して予想精度が飛躍的に向上することが判明した。またリコールなど製品に関する問題が報じられる場合は仕入れ販売関係などが関係し、財政問題に関する問題が報じられる場合は債務・債権関係が効くなどそれなりにモデルに意味が通っていることもわかった。残念ながら期間内に査読を通過することはできなかったが比較的野心的な試みであったにも関わらずプロトタイプを完成させたことは成果として大きい。

制裁リスト予想に関してはデータを整備するだけで期間が終わってしまった。しかし上研究が口コミで広がった結果、国内における制裁に関係する有識者と議論をする機会にも恵まれるなど意外な発展もあった。引き続き積極的に取り組んでいく。

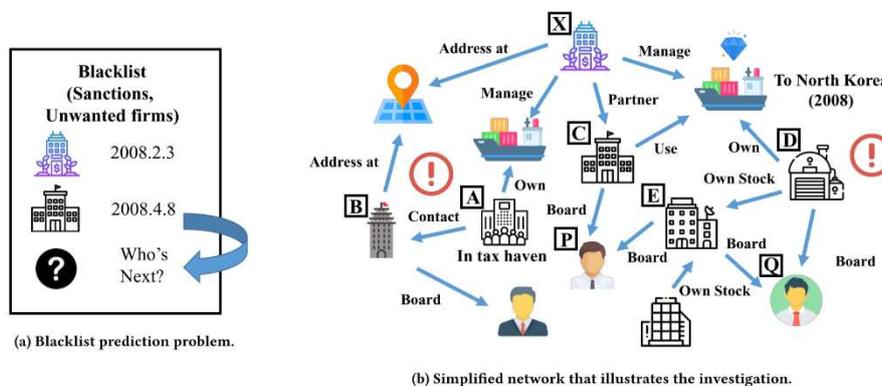


図1: (a) 問題設定 (b) 北朝鮮制裁に関する国連捜査の過程の簡略化した説明(Furukawa 2017)。詳しくは Hisano et al. (2018)参照。

(2) 詳細

研究の狙いに従い本研究ではまず企業、人、商品、株式などに関する情報を複数ソースから集め異質情報ネットワークを作成することに取り組んだ。データに関してはの表1にあるように世界中の上場企業を主な対象にした。これは投資という観点から上場企業は国際的に様々な角度からデータが整備されており、また曖昧な情報が少ないため複数ソースのデータを統合することが可能だからである。データベースの統合作業は何度かやり直す必要があったが、共同研究者や学生など複数人による利用を通じてエラーは大分少なくなったと自負している。

現行のデータベースの概要を表2と図1に示す。表2にあるように企業間の関係だけでなく商品にまつわる関係、株式にまつわる関係、人にまつわる関係など豊富に含まれていることがわかる。また図1に示している本研究で予想対象にする企業は世界中に散らばっているのも確認できる。このデータベースを用い特にテーマを ESG 投資(環境・社会・ガバナンス)に用いられる投資除外リストと制裁リストの予想に絞りモデル作成に励んだ。

ESG 投資はここ10年間関心が高まっている分野である。年金基金、保険会社などに代表される機関投資家は社会の持続的な発展に寄与するように投資することが求められており、実際、諸外国では投資の大部分が ESG 投資に相当するようになっている(図3)。日本はまだ比較的低いが国際的協調路線を取る我が国においても今後徐々に関心が高まっていくだろう。ESG 投資に関する先行研究は、(1)企業が自主的に公開した情報、(2)格付け会社が付与した情報、(3)過去のニュース情報などを用い行われてきた(Hvidkjaer 2017)。

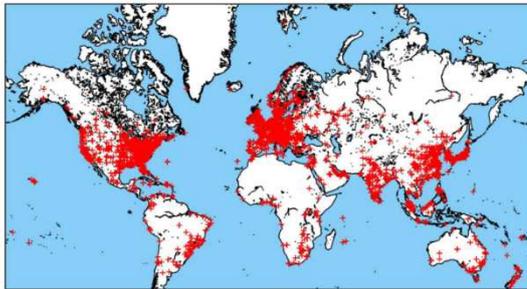
本研究では(3)のニュース情報を用い投資除外リストを作成する方法に関連した研究を実施した。従来研究の限界点は過去情報だけを用い投資除外リストを作成していることで、蓋然性としては非常に高いが単にまだ報道されていない・報道されづらい企業を無視している点にある。要するに「泥棒を捕らえて縄を絢う」状況にある。そこで本研究では投資除外リストを図1の(a)に対応させ、作成したデータベースを図1の(b)に対応させることで蓋然性を考慮できるようにした。モデルはラベル伝播法の亜種を用い複数ソースの情報を使用しないモデルに比べ予想精度が飛躍的に伸びることを示した(図3)。モデルに関してはまだまだ改善の余地があるが、最初のモデルとしては妥当な所と言える。

制裁リスト予想に関してはデータを整備するだけで残念ながら期間が終わってしまった。しかしながら Hisano et al. (2018)が口コミで広がった結果、国内における制裁に関係する有識者と議論をする機会にも恵まれており、引き続き積極的に分析していく予定である。

Source	Date of acquisition	Node types	Relation types	No. of nodes	No
Dow Jones Adverse Media Entity	Dec 2016	Firm	Location, Homepage	132,127	
Dow Jones State-Owned Companies	Dec 2016	State-owned firms	VIP, Employee, Owner	280,995	
Dow Jones Watchlist	Dec 2016	VIPs, specially interested person	Social relations	1,826,273	
Capital IQ Company Screening Report	Dec 2016	Firms	Buyer-seller, borrower etc	505,789	
FactSet	Dec 2015	Firm, goods, industry	Parent-child firm, Issue Stock	613,422	
FactShip	Jan 2017	Firm, goods, invoice etc	Overseas trade etc	16,137,550	3
Reuters Ownership	Dec 2016	Owners, stocks	Issue, own	1,560,544	12
Panama papers	Jan 2017	Entities, officers	Shareholder of, director of	888,630	
DBpedia	Apr 2016	Various	Various	35,006,127	24

表1:本研究で扱ったデータベースのソース一覧。

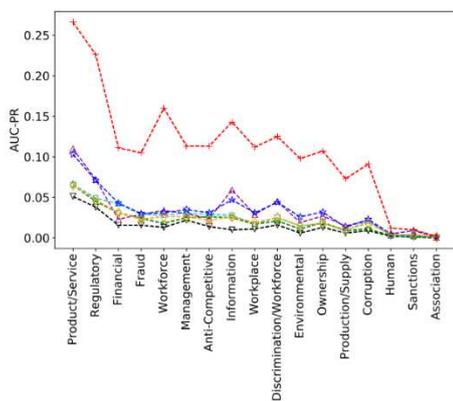
Rank	Relation	Number
1	located_in	2,723,162
2	customer	717,019
3	supplier	713,434



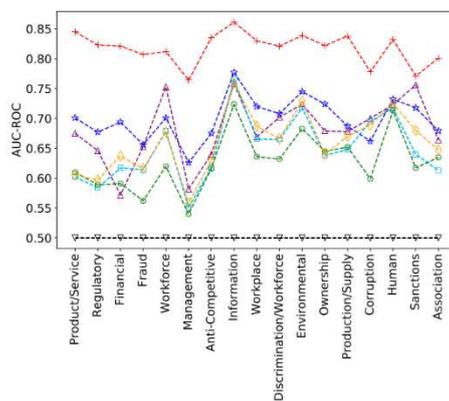
15	domain	131,153
16	distributor	116,262
17	subsidiary	107,119
18	parent-company	107,117
19	associated-person	100,699
20	international_shipping	95,050
21	associate	72,685
22	landlord	62,904
23	http://dbpedia.org/ontology/party	55,653
24	employer	47,901
25	employee	47,184

表2(左): データベースの中の出現頻度上位25の関係型。

図2(上): 予想対象企業の本社の住所を緯度経度で描画した散布図。世界中に散らばっていることがわかる。



(a) AUC-PR



(b) AUC-ROC

図3: 提案手法(赤いクロス)とその他の比較手法の比較。(a)の評価尺度は AUC-PR (b)の評価尺度は AUC - ROC。どちらも偽陽性と偽陰性の妥協度合いを一つの数字に集約した尺度である。

[参考文献]

- Furukawa, K. (2017). Kitacyosen Kaku no Shikengen Kokuren Sousa no Hiroku [Funding Source of North Korea: A Note on United Nation's Investigation]. Tokyo Shincyosya. Tokyo. Japan
- Hvidkjaer, S. (2017). ESG investing: a literature review. Report prepared for Dansif. Available at <https://dansif.dk/wp-content/uploads/2019/01/Litterature-review-UK-Sep-2017.pdf>.

3. 今後の展開

ESG 投資はここ 10 年間関心が高まっている分野である。年金基金、保険会社などに代表される機関投資家は社会の持続的な発展に寄与するように投資することが求められており (United Nations 2006; OECD 2017)、特に欧州では投資の大部分が ESG 投資に相当するようになっていく (表 3)。それに対し日本はこの潮流に出遅れているが、国際的な水準にやがては収斂していくのは間違いないだろう。

1. 狙いの繰り返しにもなるが、ESG 投資は、(1)企業が自主的に公開した情報、(2)格付け会社が付与した情報、(3)過去のニュース情報などを用い行われてきた。(1)の自主公開した情報はエンロンの粉飾決算のように偽装可能性を排除することができない。(2)の格付け情報もサブプライム問題のように利益相反により必ずしも正しく行われているとは限らない。(3)の過去ニュースはメディアの調査報道によって未知の問題を炙り出すことができる反面、蓋然的に問題を抱えているが単に報道されていない企業を察知することができないという問題を抱えている。つまり、**ESG 投資に対する関心が高まっている反面、それを支える情報環境はまだ追いついていないのである。**これは国境を越えた資本取引が年々伸びている現代社会において一層深刻な問題になっていると言える。情報環境が成熟していない中、実体を今一つ把握できない国外の企業に投資することが国際社会の持続的な発展につながるわけではないのは言うまでもない。

本研究は既に判明しているニュース情報(図 1(a))とファクト(図 1(b))を併せできる限り蓋然性の高いものは把握し国際社会の持続的な発展に関連する情報環境の改善を試みる研究である。一番の成果は従来研究がニュースが出てから反応するなど「泥棒を捕らえて縄を絢う」状況に限定されているのに対して、ある程度先の読みを可能にしたことである。同様の方法論が捜査方法にも関連することは前述の通りである。

ただ現状できているものはまだ究極の完成形には程遠いというのが正直な所である。確かに複数ソースの情報を用いる事の効果は示している。しかしまだまだモデルは改善できる。評価基準も問題がないわけではない。AUC-PR や AUC-ROC 予測器の全体的なパフォーマンスを一つの値に集約する評価基準が実務における有用性に必ずしも対応しているわけでもない。本研究の投資除外リストの裏にあるのはメディアが報道する ESG 関連ニュースである。いくらデータを整備した会社が時間をかけ選別したとはいえ、憶測で裁判沙汰になる可能性に言及している記事など深い調査報道ではなく浅薄な嫌疑によったニュースが含まれている可能性も否めない。そうした一種のフェイクニュース(定義については Zhou and Zafarani 2019 参照)を考慮した分析も重要である。同様に仮にフェイクニュースによってリストに虚偽が含まれていた場合、本研究のアプローチによる予想はどれくらい変化するだろうか? 同様の問題は制裁リスト予想も孕んでいる。間違った情報に基づき制裁リストに加えられる企業のダメージは決して小さいものではない。この点を無視している現行モデルは十分とは言えない。

こうした限界点を踏まえると本研究成果は何か「大きく解決した」というよりはパンドラの箱を開けてしまったとまとめる方が正しいのかもしれない。自分が思い描いていたビジョンが決して容易に達成できるものではないと痛感させられた 1 年半であった。

しかし ESG 投資や制裁リストの予想など十分に集められるテーマに焦点をあて具体的な研究課題を遂行したことは改めて強調に値する。ESG 投資と制裁リストのどちらも社会の持続的な発展に直接的に関係する重要な手段である。それらにおける情報環境の在り方に一石を投じた本研究は国際社会の持続的な発展に寄与することにもつながる可能性のある重要なものである。

本研究のように包括的に情報環境の在り方を研究するのは最初は ESG 投資や制裁リストなどデータが十分に集められる限定的な分野でしか実施できる研究ではないのかもしれない。しかし、例えばフェイクニュースの問題が社会に至る所で見られるように情報環境の問題は現代社会にとって広く散見される問題である。将来的にはこうした特定分野を超え広い分野に貢献ができるように今後とも一つずつ丁寧に成果を積み上げていきたい。

Region	2014	2016
Europe	59%	53%
United States	18%	22%
Canada	31%	38%
Australia/New Zealand Asia	17%	51%
Asia		1%
Japan	1%	3%
Global	30%	26%

Source: Data provided by GSIA 2014 and 2016 reports.

表3:GSIA の 2014 年と 2016 年のレポートによる ESG 投資が投資全体に占める割合(Sherwood and Pollard 2018)

[参考文献]

OECD. (2017). Responsible business conduct for institutional investors: Key considerations for due diligence under the OECD Guidelines for Multinational Enterprises. OECD guidelines (2017). <https://mneguidelines.oecd.org/RBC-for-Institutional-Investors.pdf>

Sherwood, M. W. & Pollard, J. (2018). Responsible Investing An Introduction to Environmental, Social, and Governance Investments. Routledge. ISBN 9781351361927.

United Nations. (2006). Principles for Responsible Investment. 2006. www.unpri.org. Accessed March 10, 2019.

Zhou, X. and Zafarani, R. (2019). Fake News: Fundamental Theories, Detection Strategies and Challenges. WSDM 2019. DOI:10.1145/3289600.3291382

4. 自己評価

1年半という短い期間を考えれば本研究は研究目的を達成できていると言える。データの結合作業自体は予想に時間がかかり、実際作ってみななければ何ができるかわからなかった。そのため比較的オープンエンドで研究を進め、確実に出来そうなことにテーマを絞ったことが功を奏した。次の3点:(1)当初の予定よりははるかに大きいテーマに対して具体的に応用例を完成させたこと、(2) 金融の観点からは従来の ESG 投資に関する研究がニュースが出てから反応するなど「泥棒を捕らえて縄を絢う」状況に限定されているのに対してある程度蓋然性を考慮できるようにし、一般的に問題を抱えているとされる ESG 投資にまつわる情報環境の改善に一石を投じたこと(3) コンピュータサイエンスの観点からは捜査法の分野(ACM 分類: Applied computing → Investigation techniques)に貢献したことを踏まえれば、挑戦的かつ独創的な研究であったにも関わらず一定成果を出した本研究は成功したと言える。

しかしながら反省点がないわけではない。逆にオープンエンドでやっていたからこそ最初自分自身も研究成果がどういう意味を持つのか全体をなかなか把握しきれずコミュニケーションがうまく

いかなかった。スケジュールに関しても何度かデータベース作成からやり直さなければいけなかったこともあり、一人の人間が実施する研究としてはぎりぎりの作業量だったのかもしれない。今後はこうしたスケジュール管理もうまく考慮できるようにしていきたい。

研究費用で購入したサーバーとデータに関しては有効活用させていただいている。計算資源が増えたおかげで並列でアルゴリズムを走らせられるようになったおかげで研究の進みが速くなったのは言うまでもない。

5. 主な研究成果リスト

(1) 論文(原著論文) 発表

1. Hisano, R., Sornette, D., and Mizuno, T. (2018). Social Blacklist Prediction using a Heterogeneous Information Network. <https://arxiv.org/abs/1811.12166>. Submitted.

(2) 特許出願

研究期間累積件数: 0 件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

[学会発表]

2018年12月4日、Predicting Adverse Media Risk using a Heterogeneous Information Network, キヤノングローバル戦略研究所(CIGS): 経済・社会の分野的横断研究会

2019年1月22日、Predicting News Events using a Heterogeneous Information Network, 3rd ETH Zürich-UTokyo Strategic Partnership Symposium on the United Nations Sustainable Development Goals and Innovation