

社会課題解決のための ビッグデータ処理と利活用事例

水野貴之
国立情報学研究所

自己紹介

水野貴之

国立情報学研究所 情報社会相関研究系 准教授

総合研究大学院大学複合科学研究科 准教授

学位: 博士(理学)

専門分野: 計算社会科学、データサイエンス

【研究背景】

格差社会・バブル・テロなど、グローバル社会が抱えるあらゆる問題は、それぞれが関係しあい、影響をあたえ、複雑な繋がりを形成しています。

【研究テーマ】

グローバル社会を複雑ネットワーク解析で紐解き
グローバル問題を経済の力で解決するAI

- グローバル経済ネットワークの安全保障
- バブル、金融危機
- 排他的民族主義、富の不平等
- SNSによる世論操作

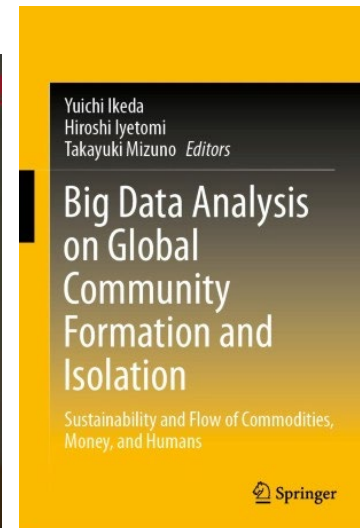
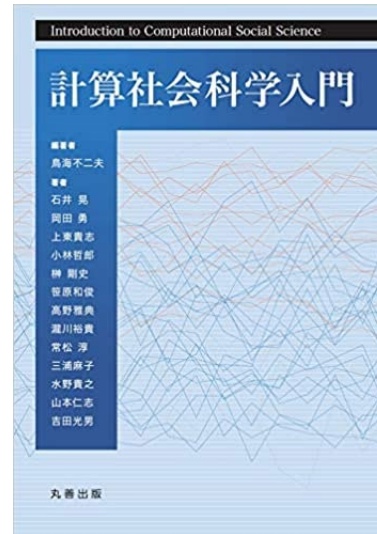
【受賞】

人工知能学会 2022年度全国大会優秀賞

人工知能学会 2021年度全国大会優秀賞

人工知能学会 2019年度全国大会優秀賞

情報処理学会第81回全国大会大会優秀賞



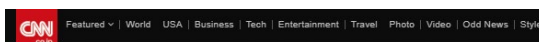
目次

1. **社会課題と数理モデル: グローバル経済システムの安全保障**
 - 経済優先・性善説で何でも公開すればよいというものではない
 - 数理とビッグデータから隠れた支配者を発見する
2. 情報処理: 深層学習によるデータ間の接合
 - 土地安全保障, サプライチェーン安全保障
3. 利活用: PoC, フィジビリティスタディ
 - 経済安全保障可視化システム NPI Visualization
4. 情報公開: オープンデータ化のハードル
 - シンセティックデータ生成

経済安全保障問題 – 隠れた外国政府の影響 –

水野, 土井, 栗崎(2023)人工知能学会誌 38(2), 101-108.

- ① 「資本関係(+株式市場)」
 - 上場企業に関して資本上で外国政府の影響がないこと
- ② 「資本関係+役員」
 - 外国政府の影響のある人物を雇用してしまわないか
- ③ 「資本関係+サプライチェーン」
 - 調達/販売経路において外国政府や外国軍の影響がないこと
- ④ 「資本関係+登記簿」
 - 重要施設周辺の地権者に外国政府の影響がないこと



フィリピンの電力網、中国が「いつでも遮断可能」 内部報告書が警告

© 2019, 11:26 Tue posted at 17:58 JST



① 資本

アプリダウンロード禁止
海外上場規制強化

2020年12月成立
外国企業説明責任法

ティティ 滴滴

BOSS 直聘

上場廃止

プーチン大統領

シロビキ 最接近 (元KGBなど)

② 人物

オリガルヒ(約250人)
プーチン政権の元で
富を築いてきた財閥の人々

新疆ウイグル自治区

強制労働疑い

不買運動

人権侵害なし確認できるか

輸入する場合は
強制労働でない証明
提出を!!

アメリカとの
ビジネスにも影響?

③ サプライチェーン

半導体

レアメタル・レアアース

医薬品 など

特定重要物資に指定し
調達ルートや保管状況を国がチェック

注視区域

●重要施設の周辺 自衛隊・米軍基地、原発など
●国境離島等

④ 土地

特別注視区域

●特定重要施設 司令部機能のある基地など
●特定国境離島等

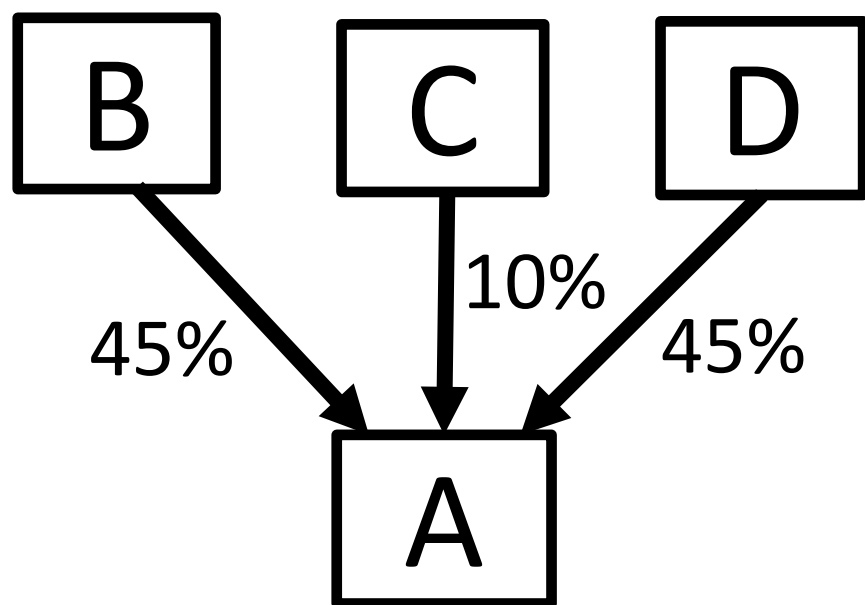
Network Power Index = 潜在的な意思決定の伝搬過程

- ◇ アスクル株式会社
- ◇ (親会社) ヤフー株式会社
- ◇ (祖父母会社) ソフトバンク株式会社
- ◇ (曾祖父母会社) ソフトバンクグループジャパン株式会社
- ◇ (高祖父母会社) ソフトバンクグループ株式会社
- ◇ (親の5乗) 孫正義

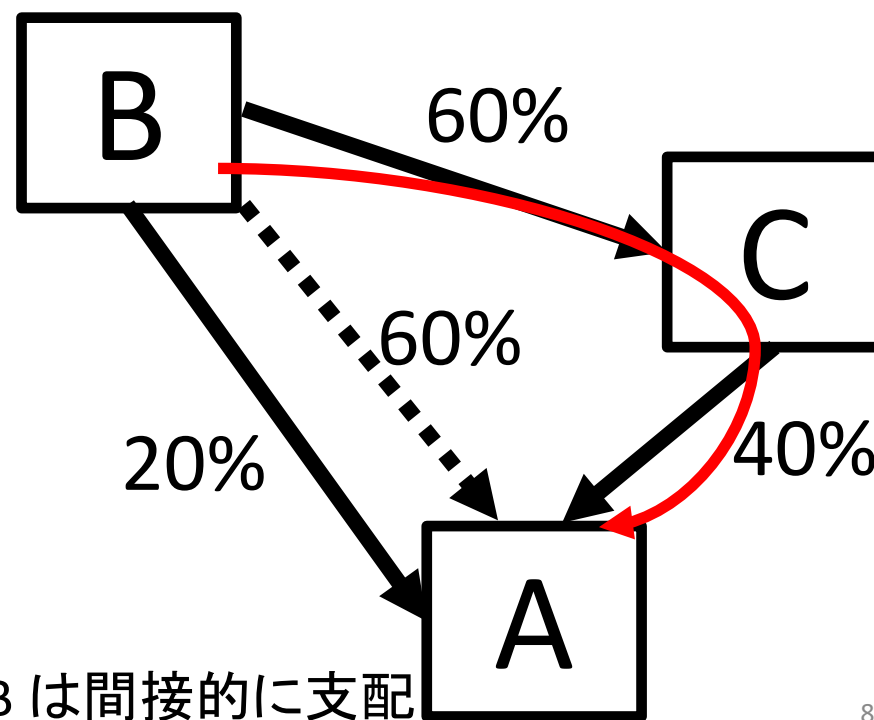


協力ゲーム（投票ゲーム）

- 民主主義ルール「多数決」で誰がキャスティングボードを握るか？
例：議席数7~10%でも公明党の発言力は強い
- 資本関係における各ステークホルダーの影響力を理論的に算出
つまり、隠れ支配者が分かる

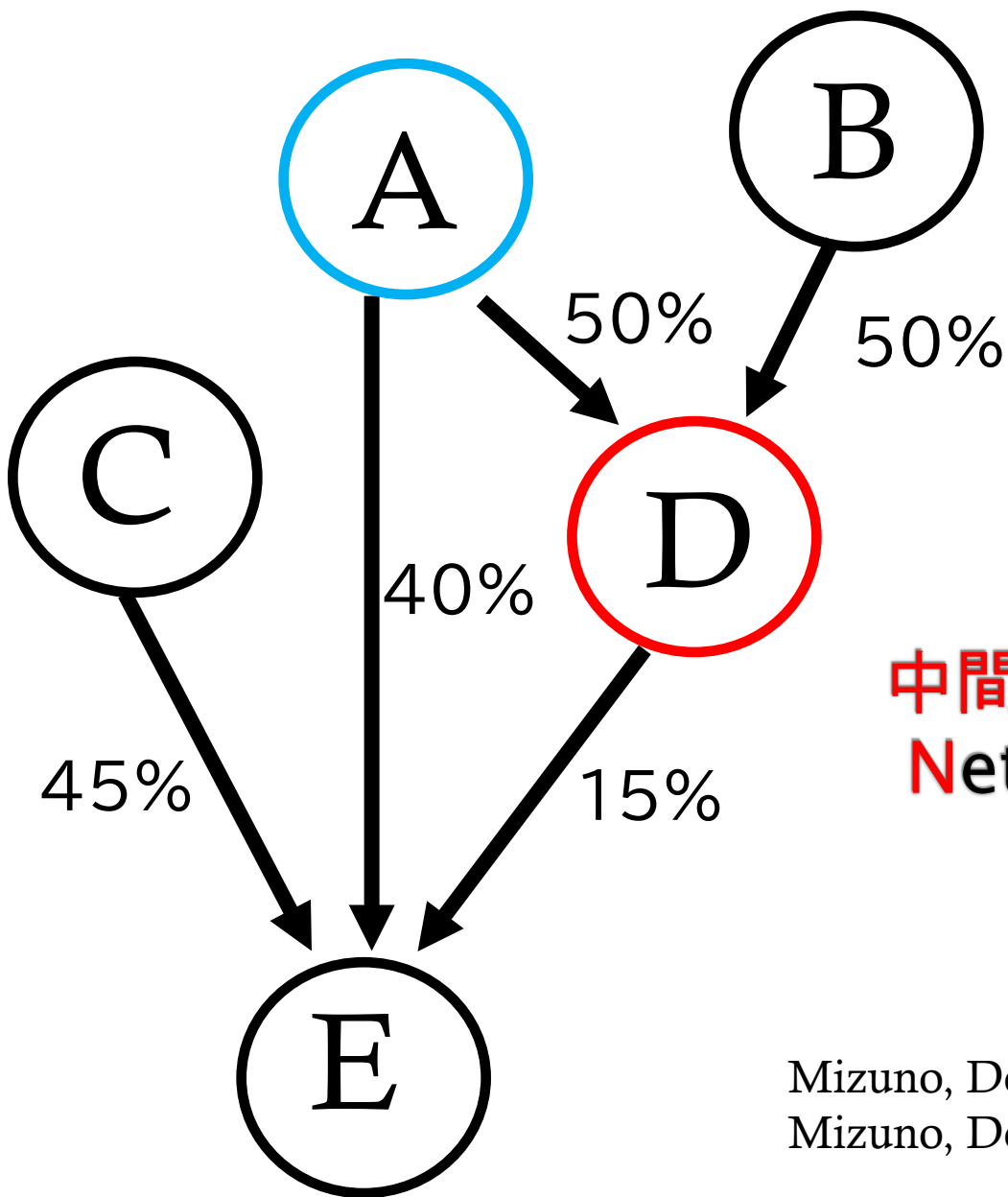


株主 C がいないと可決できない！



株主 B は間接的に支配

影響力(意思決定)の流れの数理モデル



ネットワーク効果を加味したShapley-Shubik power indexと、それに更にPageRank(流れ)の概念を加えた指数を導入

最終株主の影響力

Network Power Index

$$NPI(A \rightarrow E) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot 1 = \frac{2}{3}$$

中間株主の影響の仲介力

Network Power Flow

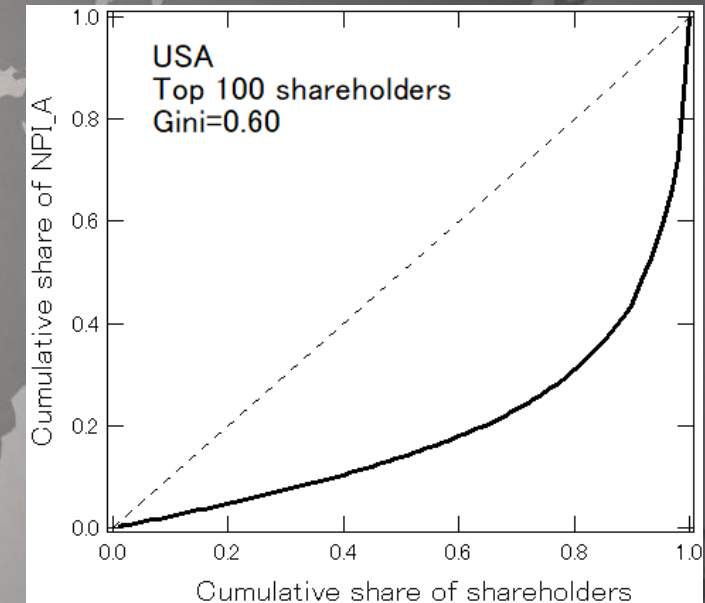
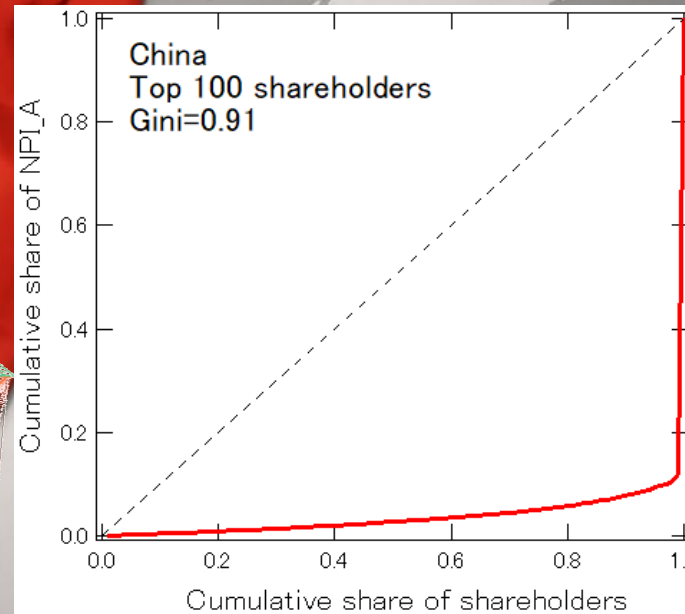
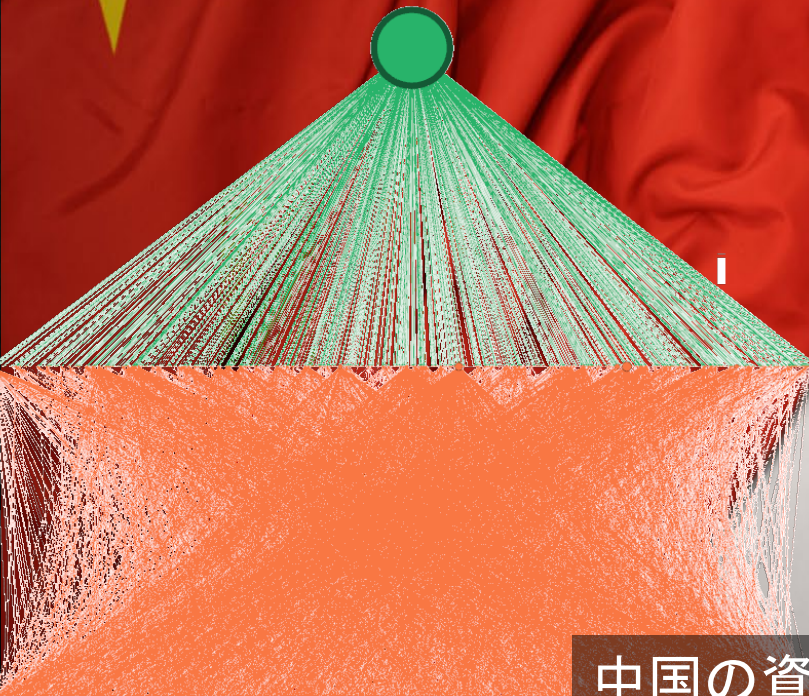
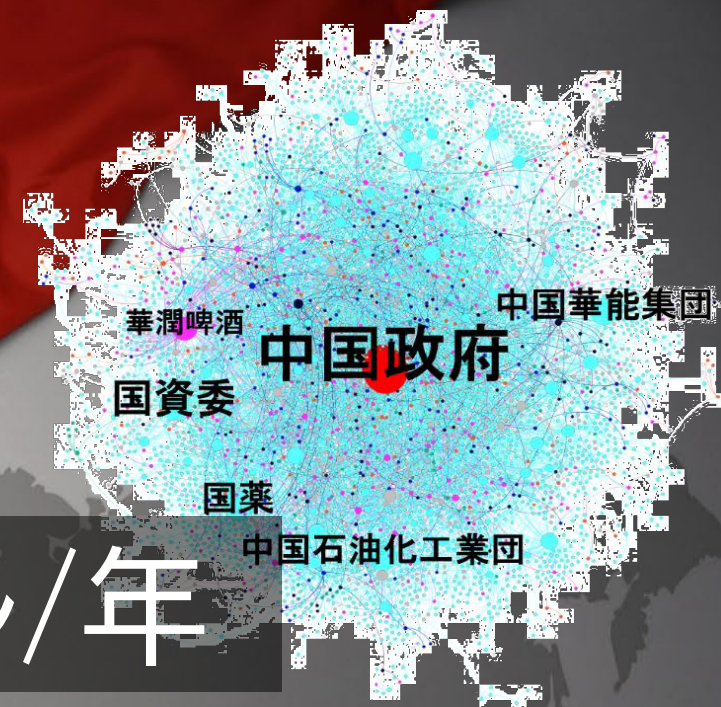
$$NPF(D \rightarrow E) = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

Mizuno, Doi, Kurizaki (2020) PLoS ONE 15(8), e0237862.

Mizuno, Doi, Tsuchiya, Kurizaki (2021) PLoS ONE 16(8), e0256160

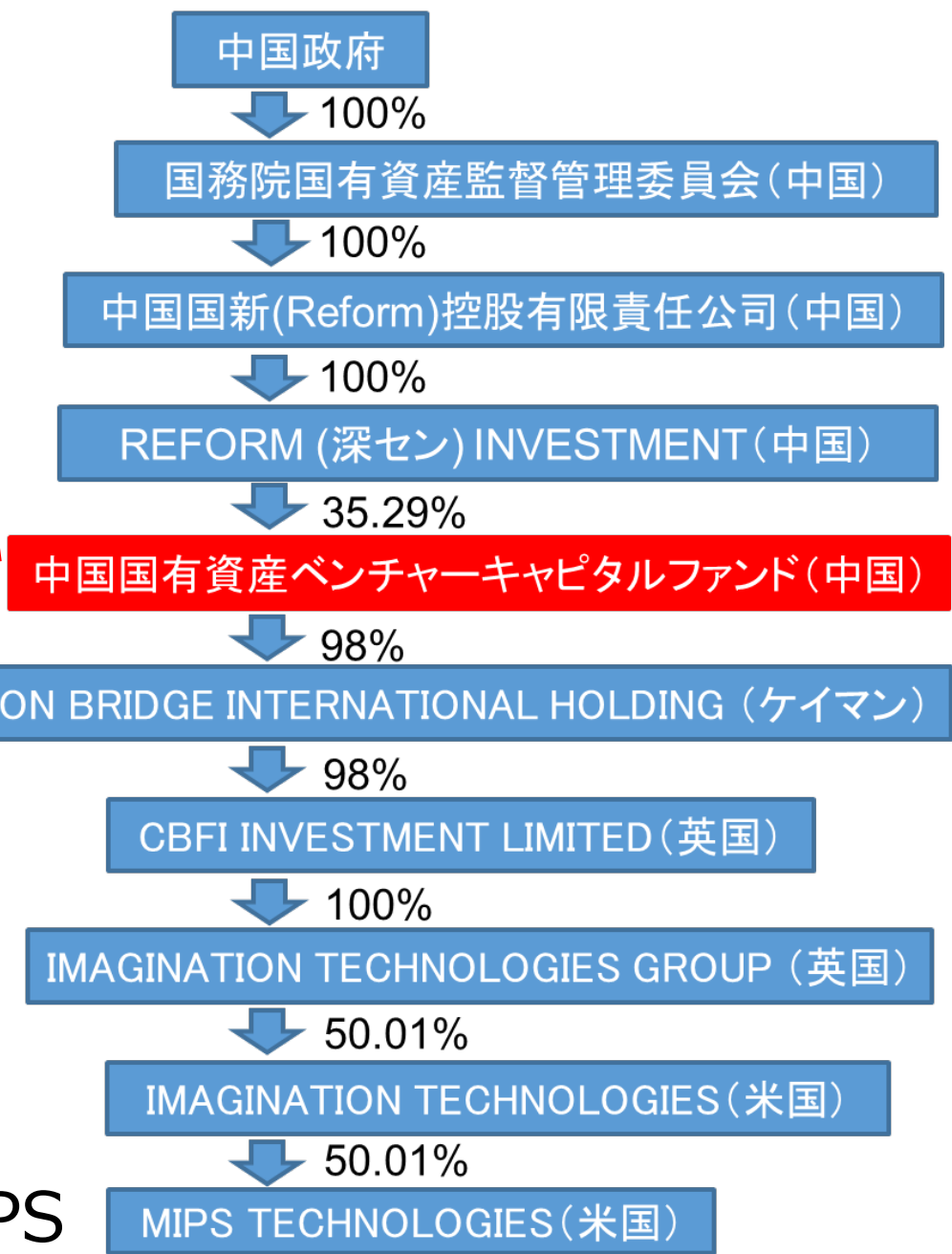
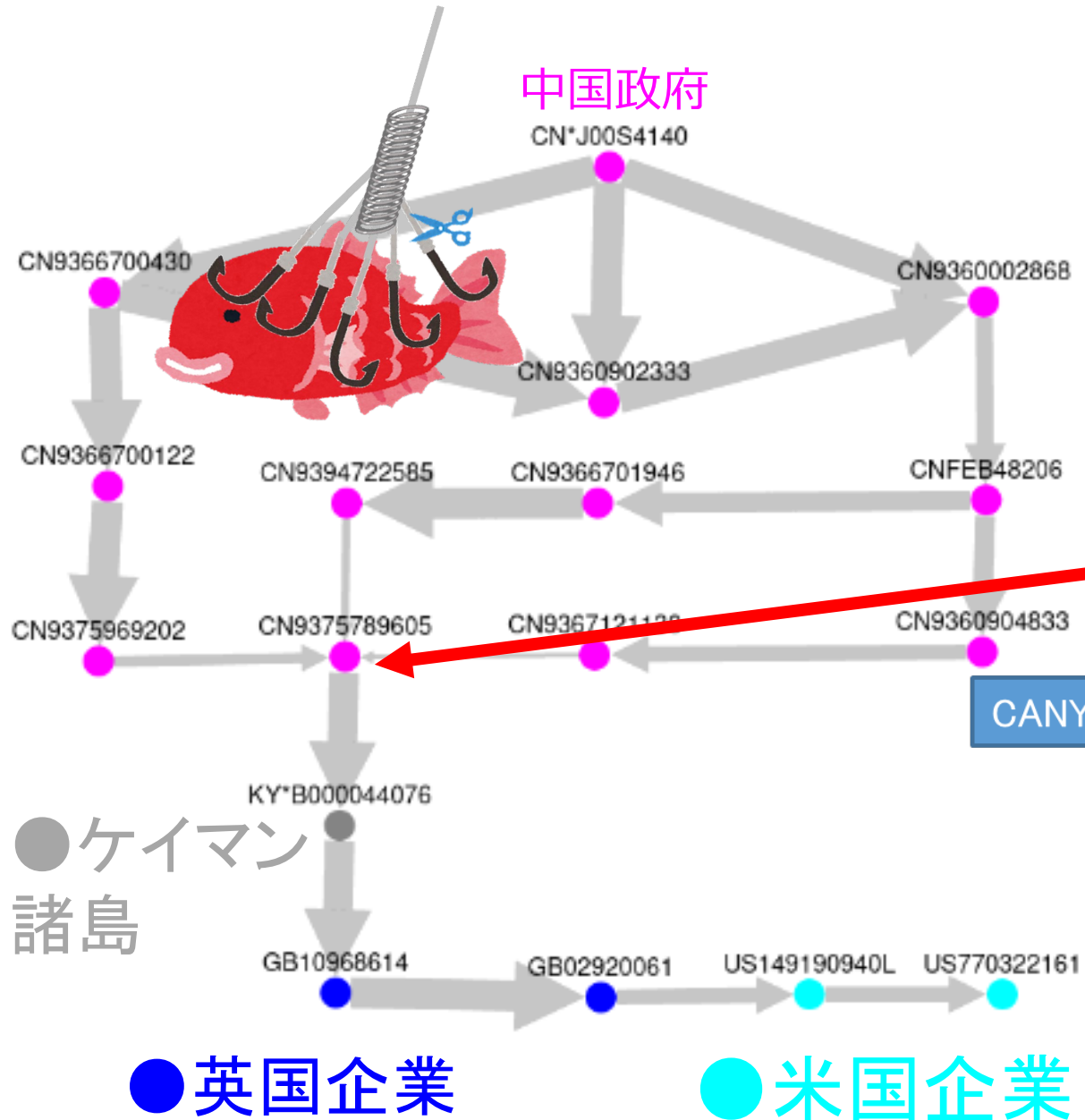
4億中1位

$NPI = 15.21$ 兆ドル/年



中国の資本関係はピラミッド構造
すべての力が芽づる式に中国政府に集約する

資本構造での外国政府の多面的な影響力の見える化



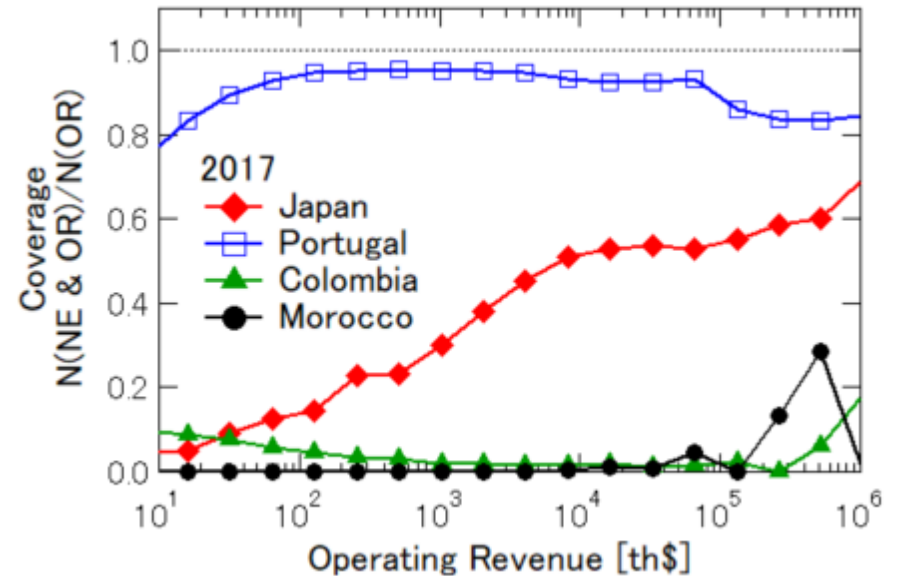
米国 半導体メーカーMIPS

非ランダム欠損

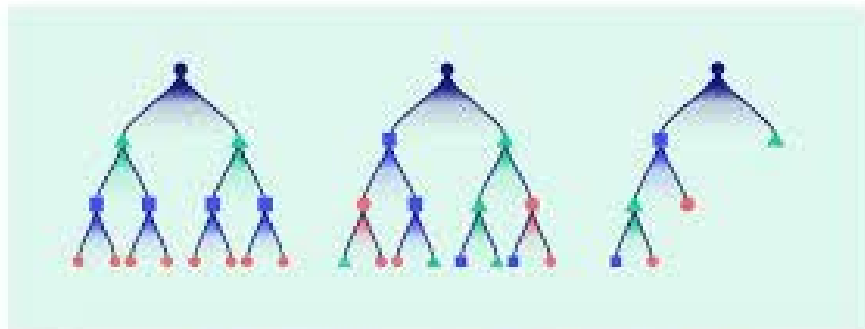
S. Fujimoto, T. Mizuno, A. Ishikawa (2022) Journal of Computational Social Science 5, 1281-1301.

- 需要のない(中小企業などの)財務項目には欠損が多い
- 需要のない財務項目には欠損が多い

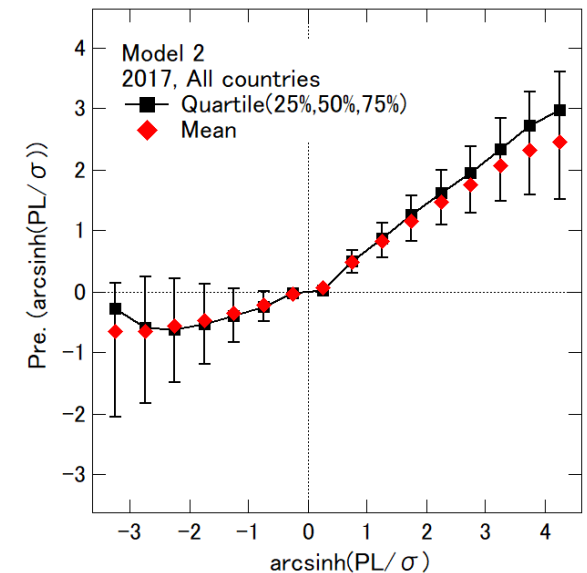
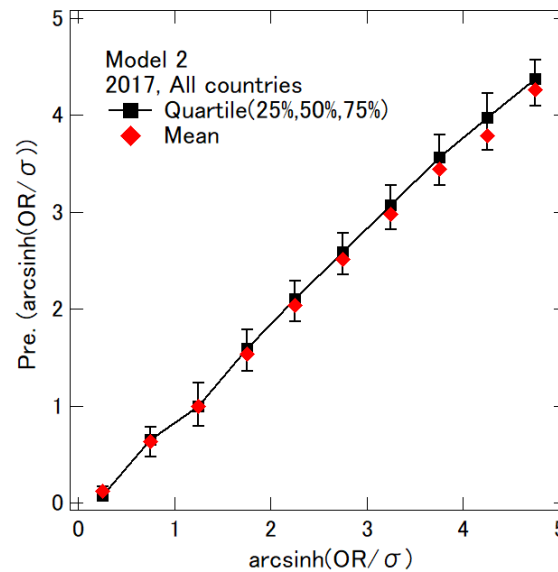
各国企業における
売上規模別の従業員数収録率



機械学習ブースティング・ツリー
Catboostで補完



CatBoost - An In-Depth Guide
[Python]



目次

1. 社会課題と数理モデル: グローバル経済システムの安全保障
 - 経済優先・性善説で何でも公開すればよいというものではない
 - 数理とビッグデータから隠れた支配者を発見する
2. 情報処理: 深層学習によるデータ間の接合
 - 土地安全保障, サプライチェーン安全保障
3. 利活用: PoC, フィジビリティスタディ
 - 経済安全保障可視化システムNPI Visualization
4. 情報公開: オープンデータ化のハードル
 - シンセティックデータ生成

目次

1. 社会課題と数理モデル: グローバル経済システムの安全保障
 - 経済優先・性善説で何でも公開すればよいというものではない
 - 数理とビッグデータから隠れた支配者を発見する
2. 情報処理: 深層学習によるデータ間の接合
 - 土地安全保障, サプライチェーン安全保障
3. 利活用: PoC, フィジビリティスタディ
 - 経済安全保障可視化システム NPI Visualization
4. 情報公開: オープンデータ化のハードル
 - シンセティックデータ生成

影響力評価システム NPI VISUALIZATION

Variety Velocity Volume Visual Value



水野貴之, 土井翔平: 影響力評価システム, 特願2019-119796

水野貴之, 土井翔平, 栗崎周平: 影響力評価システム, 特願2021-162219

水野貴之, 土井翔平: 情報検索システム, 特願2022-030948

英国企業の登記簿

Dataset

UK companies that own property in England and Wales



ピーク・ディストリクト国立公園

ミネラルを含んだ
天然水が豊富

多くの表記ゆれやスペルミスが存在

Contents

- [The dataset](#)
- [Get example data](#)
- [Access the data](#)
- [Contact us](#)

The dataset

This data shows registered land
by UK companies

Inconsistencies in the data

The data comes from different sources. This means that not everything is written in the same way. For example, one address may say '24 The Willows' and another might be written as 'Field south of N14'.

You may also see:

- spelling errors or typos
- descriptions instead of addresses (land being developed may not have an address)
- partial postcodes or no postcodes at all
- mistakes in company registration numbers
- no data for date of registration
- 2 different companies with the same registration number

デモ動画

(英国) 不動産の所有者への外国政府の影響

指定座標の周辺の不動産と、その所有者

緯度と経度、検索範囲を入力してください。

例 [53.33375023252062][-1.7846968240574235]:ピーク・ディストリクト国立公園, [53.60518760958058][-1.8940490216670693]:ディー・ヒル貯水池

例 [3000] : 3000m=3km

緯度: 経度:

検索範囲 (半径): m (メートル)

土地リスク検索

座標: ピーク・ディストリクト
国立公園

範囲: 30km圏内

原発や軍用地周辺の土地の調査可能

地権者への外国政府の影響の見える化


水野, 土井, 栗崎(2023)人工知能学会誌 38(2), 101-108.

ヨークシャー・ウォーター株式会社

ケルダグループ

シンガポール政府

NPI=0.33



YorkshireWater

タイプ プライベート
 業界 水産業
 設立 1973年
 本部 ブラッドフォード, イングランド
 サービスエリア ヌークシャー
 重要人物 委員長 - ヴァンダ・マレー (2021) [1]
 CEO - ニコラ・ショー(2022) [2]
 製品 水を飲んでいる
 リサイクルされた廃水
 生産量 1.3GI /日 (飲用)
 1.0GI/日 (リサイクル)
 収益 11 億ポンド (2020 ~ 2021 年) [3]
 就業者数 2,500
 親 ケルダグループ
 Webサイト www.yorkshirewater.com _
 _[🔗](#)

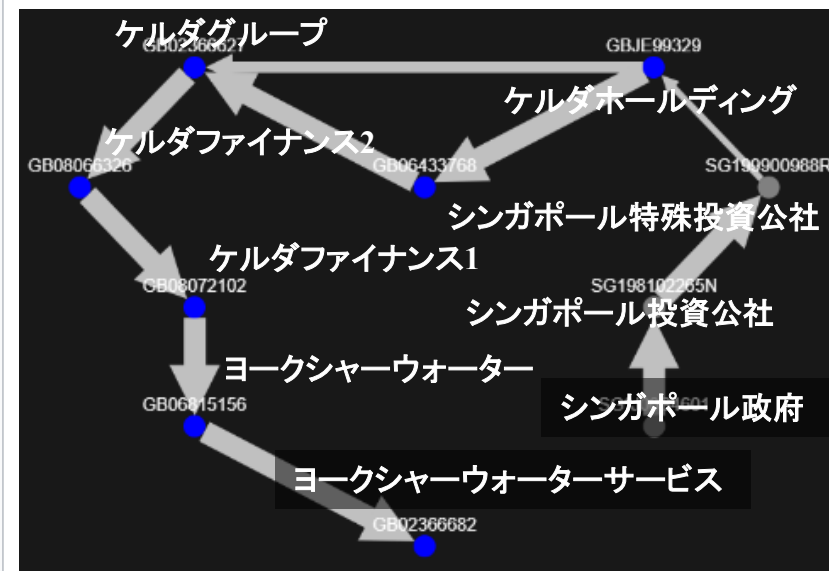
100%




KeldaGroup

タイプ プライベート
 業界 効用
 設立 1989年
 本部 ブラッドフォード, イングランド, イギリス
 重要人物 アンソニー・ラビン (議長) [要出典]
 Nicola Shaw (CEO) [要出典]
 収益 8億7890万ポンド (2007年)
 営業利益 3億3820万ポンド (2007年)
 当期純利益 1億5,130万ポンド (2007年)
 オーナー ソルテアウォーター
 就業者数 3,843人 (2007年)
 子会社 ヌークシャー水道、ケルダ水道
 Webサイト www.keldagroup.com [🔗](#)

33.56%



デモ動画

サプライチェーン・リスク検索

検索フォーム

検索ワード例

企業 [TOSHIBA CORPORATION], [TESLA, INC.], [MIPS TECHNOLOGIES INC]

調達財 [lithium battery], [cotton], [gold,tantalum,tungsten,tin]

企業名: 調達財 (部品など):

調達経路のリスク検索

企業: テスラ

部品: バッテリー

「外国政府からの関与」

Elkem ASA

「温室効果ガスへの関与」

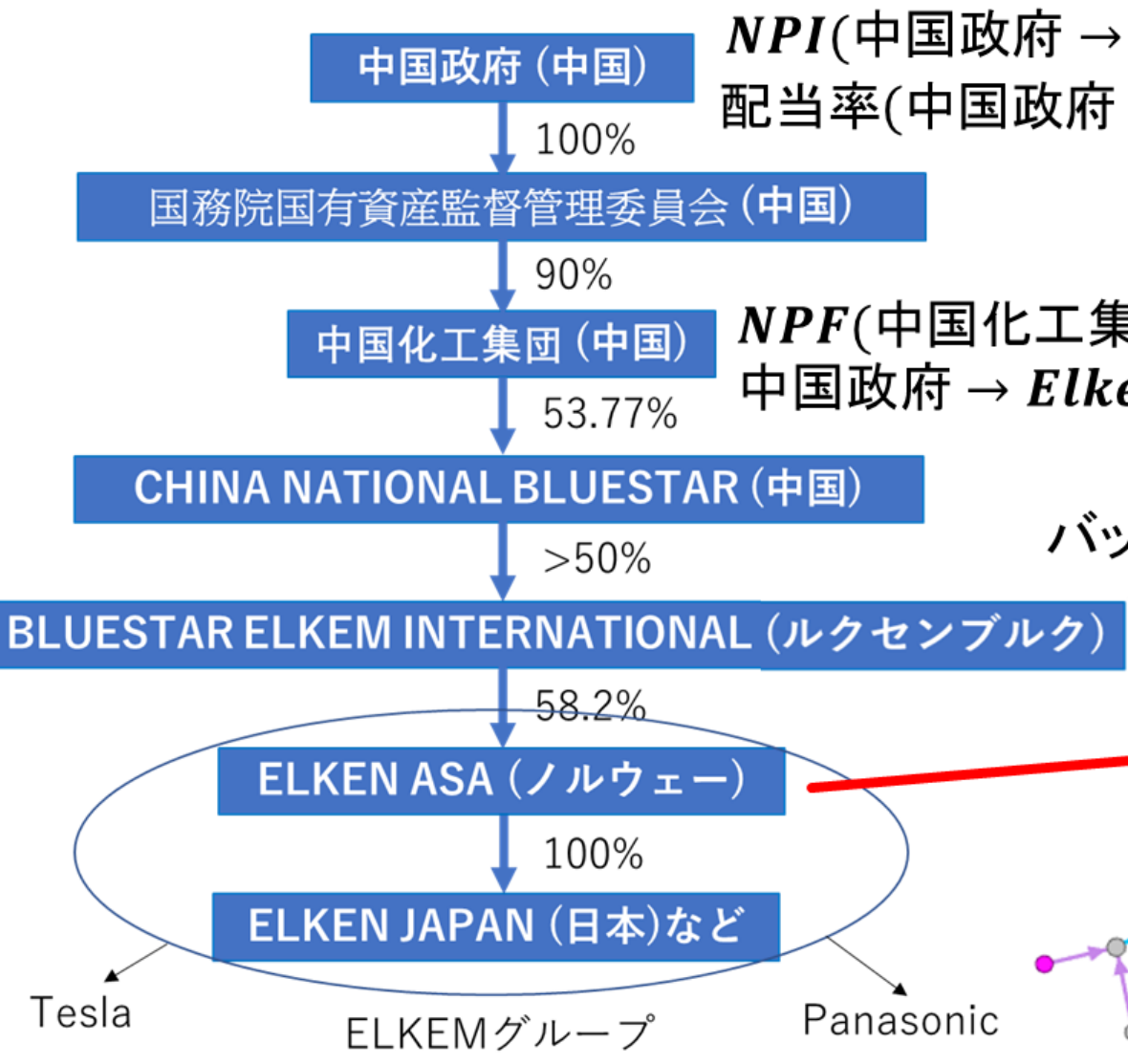
Mechel PAO

POSCO

調達経路での外国政府の影響の見える化

栗崎(2023.1.6)日経ビジネス

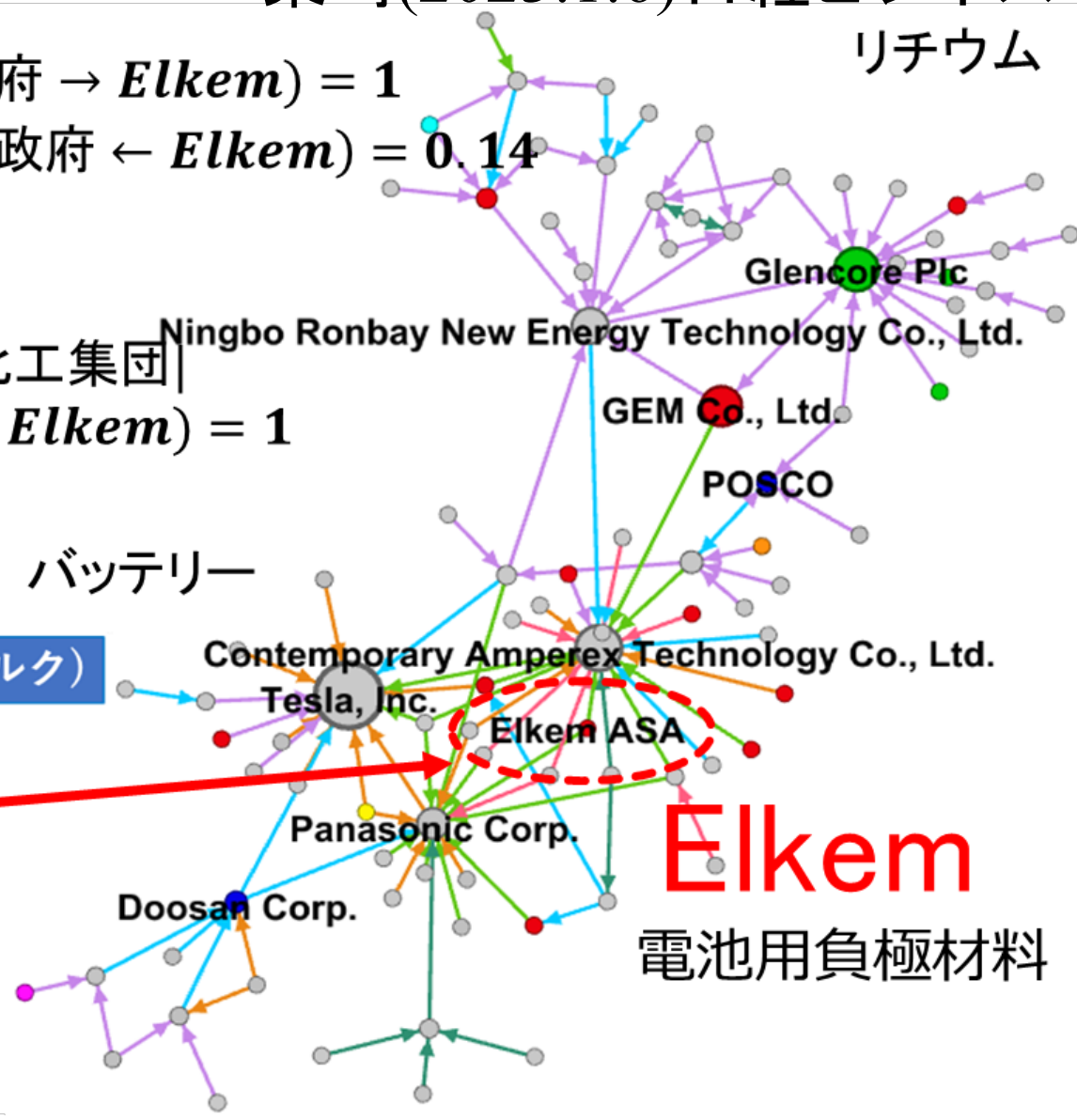
リチウム



$NPI(\text{中国政府} \rightarrow \text{Elkem}) = 1$
 $\text{配当率}(\text{中国政府} \leftarrow \text{Elkem}) = 0.14$

$NPF(\text{中国化工集団} | \text{中国政府} \rightarrow \text{Elkem}) = 1$

バッテリー



テスラ(電気自動車)の統計的に推定されたバッテリー・サプライチェーン

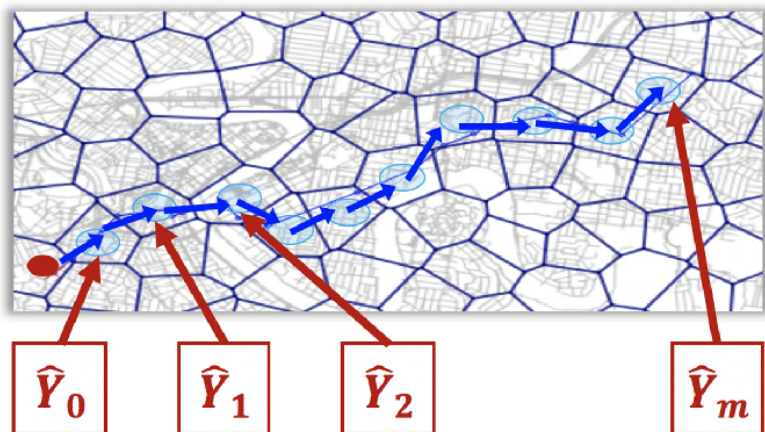
目次

1. 社会課題と数理モデル: グローバル経済システムの安全保障
 - 経済優先・性善説で何でも公開すればよいというものではない
 - 数理とビッグデータから隠れた支配者を発見する
2. 情報処理: 深層学習によるデータ間の接合
 - 土地安全保障, サプライチェーン安全保障
3. 利活用: PoC, フィジビリティスタディ
 - 経済安全保障可視化システムNPI Visualization
4. 情報公開: オープンデータ化のハードル
 - シンセティックデータ生成

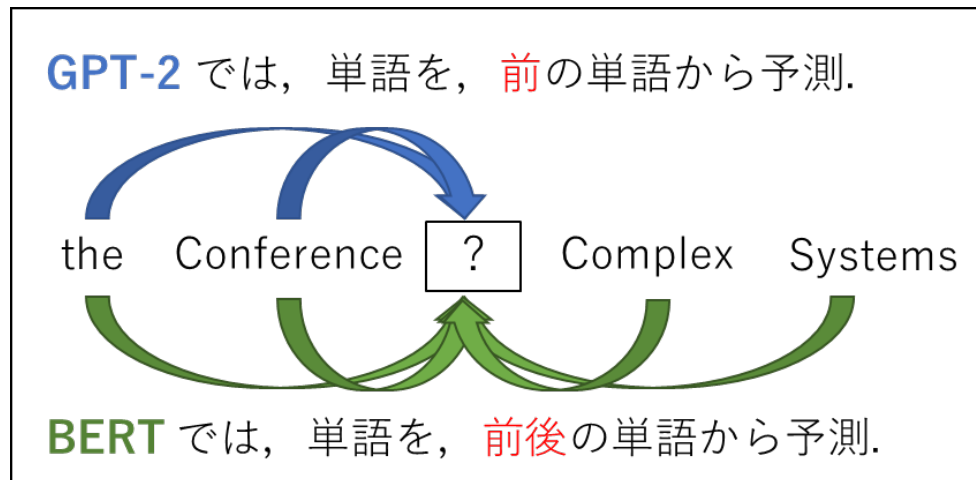
GPT-2によるシンセティック移動データ生成

T. Mizuno, S. Fujimoto, A. Ishikawa (2022) Front. Phys. 10:1021176

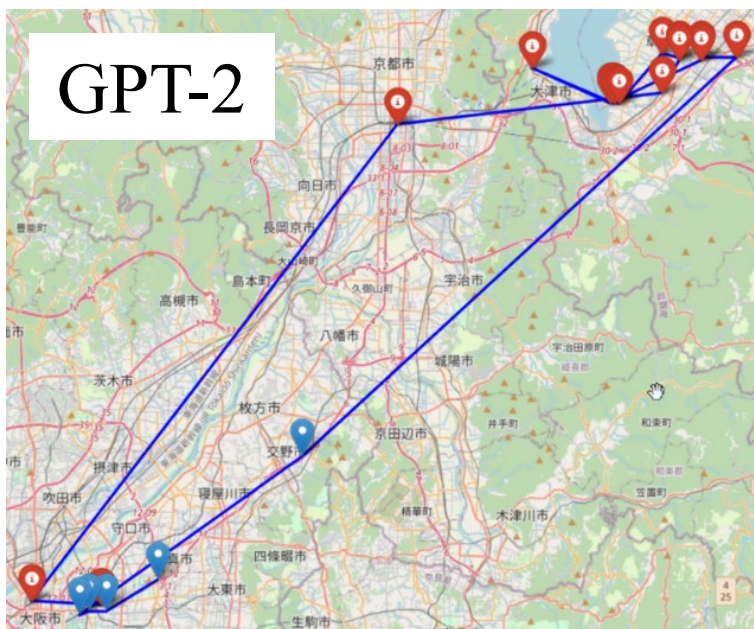
移動時系列



文字時系列 = 文章



GPT-2



正解の1km範囲, 10km範囲に入る確率

	30分後	2時間後	4時間後	1日の最後
1次マルコフ	0.29 (0.67)	0.073 (0.42)	0.035 (0.33)	0.011 (0.20)
2次マルコフ	0.33 (0.75)	0.11 (0.46)	0.053 (0.33)	0.014 (0.19)
Catboost	0.15 (0.70)	0.041 (0.45)	0.034 (0.40)	0.016 (0.25)
GPT-2	0.40 (0.82)	0.16 (0.55)	0.10 (0.43)	0.12 (0.40)

まとめと社会実装への展望

- 経済以外にも安全保障への活用がある
- 民間等のビッグデータには欠損が多い
- ビッグデータの接合に課題がある
 - 古典的な欠損処理や接合処理ではなく、深層学習を利用しよう
- 近年は、基礎研究から社会実装まで求められている
 - Webアプリケーションの開発
 - 官庁や企業とのPoC, フィジビリティスタディ
- コスト面とセキュリティ面のハードルが上昇
 - シンセティックデータの公開が1つの解決策である

本研究成果の一部は、既にプロトタイプを開発し省庁や企業の協力によりPoCを終えています。研究成果を広く社会実装するため、現在、企業による製品・サービス化の提案を募集しております。

ご感心ある方は → mizuno@nii.ac.jp まで