E-028

新聞コーパスを用いた SDGs が掲げるターゲットにおける話題性の時系列トレンド
Time-series trends in topicality of SDG targets using newspaper corpus

毛利 研† 春日 剛‡ 大場 久永† 中原 えりか‡
Ken Mohri Takeshi Kasuga Hisanaga Oba Erika Nakahara

1. はじめに

機械学習の技術が近年急速に発展したことにより，ビッグデータを利用したモデリングの信頼性が向上し，かつ容易になった．その結果，金融業界内でアルゴリズム取引の導入が進み，銘柄選択や投資比率の変更，新たな投資戦略の導入等によって追加的に得られるリターン生成（アルファ値）のための新たな情報源としてオルタナティブデータの利用が積極的に行われている[1].

オルタナティブデータとは，投資家によって投資判断のために使われるデータのうち，伝統的に用いられてきた決算開示を含む一般的な公開情報以外のデータを指す．例えば，ニュースの記事，SNS の投稿，POS データ，クレジットカード決済情報，気象情報，人工衛星からの画像情報などがあり，従来投資判断に使うことが難しかったデータが，海外の金融機関を中心に利用が広まっている[2]．その中でも，近年，持続可能な開発目標（SDGs：Sustainable Development Goals）に対する投資を加速させるために必要なデータなどの定義や提供が求められている．

SDGs とは，2001 年に策定されたミレニアム開発目標（MDGs）の後継として，2015 年 9 月の国連サミットで加盟国の全会一致で採択された「持続可能な開発のための 2030 アジェンダ」に記載された，2030 年までに持続可能でよりよい世界を目指す国際目標である[3]．合計 17 のゴール・169 のターゲットから構成され，地球上の「誰一人取り残さない（leave no one behind）」ことを誓っている．しかし，SDGs が掲げる世界の貧困や飢饉などの解決には長期的かつ巨大な資金が必要だが，気候変動や人口動態の変化，世界市場の変化などによる影響を定量化したり比較したりするための情報の整備が初期段階に留まっているため，企業などが投資する際の判断材料が不十分である．そのため，SDGs 関連分野を対象にした ESG（Environment（環境），Social（社会），Governance（企業統治）投資の拡大に向け，ESG データへのアクセスと，ESG に関するオルタナティブデータが必要だと言える．

本研究において過去約 10 年分の新聞記事コーパスデータを用いて，SDGs に対するオルタナティブ・データ活用を検討するべく，自然言語処理を用いて SDGs が掲げるターゲットにおける話題性の時系列トレンドや注目するべきセクタの抽出を試みた．

2. 研究のアプローチ

経済の変動は経済指標や株価に代表されるように経済の変動はファンダメンタルズと投資家の投資行動における心理的な側面の両方が影響し，多数の市場参加者の膨大な情報が相互に作用しあって決定されるため，膨大な情報が相互に作用しあって決定される．非線形な挙動を示すため，数値情報の分析結果のみで長期的な投資判断を下す事は難しいとされている．その理由として，人間が解釈を行えないほど複雑な最適化が生じることや，数値以外の情報が欠落してしまっていることが挙げられる．そこで正しい用字用語で事象を的確に伝え，しかも裏付けのあるファクト（事実）を発信している新聞記事のテキストを用いて経済の変動を分析する研究が，自然言語処理技術の発達と共に近年盛んになってきている[3].

本研究では，上記を受けて日本経済新聞社が提供する日経コーパスを用いた[4]．データにはタイトル，本文のほか，業界，会社名，人物名，キーワードなど詳細情報のタグを付与してあり，分析することが容易なコーパスとなっていることが特徴である．なお，テキストを用いる利点として解釈が容易であること，数値に含まれない情報も分析できることの 2 点が挙げられる．

SDGs が掲げるターゲットにおける話題性の時系列トレンドは，この取得した記事の特徴ベクトルと SDGs ターゲットの特徴ベクトルのコサイン類似度を計算，類似度が高い記事の件数の推移を算出することで可視化した．

2.1 SDGs の目標とターゲット

本研究では，SDGs のエネルギーに関する目標 7（表 2.1-1 参照）および気候変動に関わる目標 13（表 2.1-2 参照）に注目する．それは，ESG 投資家と呼ばれる巨大な機関投資家を中心に，金融機関による気候変動防止のための投融資が積極的に行われているからである．また，世界の政治・経済界の指導者が参加する世界経済フォーラム（WEF）は，年次総会（ダボス会議）の開催に際して世界がその年に直面するリスクをまとめた「グローバルリスク報告書」を発表 [5]，2020 年には「異常気象」「気候変動対策の失敗」といった環境にかかわるリスクが上位を占めた．したがって，産業界がどのようにこれら SDGs が掲げられる前から過去約 10 年に渡って関わって来たのかについて明らかにする．

2.2 新聞記事コーパス

使用した日経コーパスは，2011 年 6 月 1 日から 2019 年 12 月 31 日までの約 10 年間分である．各年の本文記事数は，図 2.2-1 に示す通りであり，合計 746,001 件分析に使用する．

データクレンジング は，「▽●■▲▼≪≫」に代表される特殊文字，URL や Copyright を削除，また，アルファベットの大文字を小文字化，半角文字を全角文字に変換，文字コードの統一を図る．本文記事の平均文字数は 582 文字であり，分析で使用する記事として 300 文字以下の記事は除いてある．また，記事と SDGs のターゲットで構成されるキーワードとの演算を行うため文単位で区切り，分かち書きをする．

† 有限責任監査法人トーマツ デロイトアナリティクス
Deloitte Touche Tohmatsu LLC, Deloitte Analytics
‡ 株式会社国際協力銀行
Japan Bank for International Cooperation

表 2.1-1 SDGs 目標 7 におけるターゲット

7.	エネルギーをみんなに そしてクリーンに すべての人々に手ごろで信頼でき，持続可能かつ近代的なエネルギーへのアクセスを確保する
7.1	2030 年までに，安価かつ信頼できる現代的エネルギーサービスへの普遍的アクセスを確保する．
7.2	2030 年までに，世界のエネルギーミックスにおける再生可能エネルギーの割合を大幅に拡大させる．
7.3	2030 年までに，世界全体のエネルギー効率の改善率を倍増させる．
7.a	2030 年までに，再生可能エネルギー，エネルギー効率及び先進的かつ環境負荷の低い化石燃料技術などのクリーンエネルギーの研究及び技術へのアクセスを促進するための国際協力を強化し，エネルギー関連インフラとクリーンエネルギー技術への投資を促進する．
7.b	2030 年までに，各々の支援プログラムに沿って開発途上国，特に後発開発途上国及び小島嶼開発途上国，内陸開発途上国の全ての人々に現代的で持続可能なエネルギーサービスを供給できるよう，インフラ拡大と技術向上を行う．

表 2.1-2 SDGs 目標 13 におけるターゲット

13.	気候変動に具体的な対策を すべての国々において，気候変動に起因する危険や自然災害に対するレジリエンスおよび適応力を強化する．
13.1	すべての国々において，気候変動に起因する危険や自然災害に対するレジリエンスおよび適応力を強化する．
13.2	気候変動対策を国別の政策，戦略および計画に盛り込む．
13.3	気候変動の緩和，適応，影響軽減，および早期警告に関する教育，啓発，人的能力および制度機能を改善する．
13.a	重要な緩和行動や実施における透明性確保に関する開発途上国のニーズに対応するため，2020 年までにあらゆる供給源から年間 1,000 億ドルを共同動員するという，UNFCCC の先進締約国によりコミットメントを実施し，可能な限り速やかに資本を投下してグリーン気候基金を本格始動させる．
13.b	女性，若者，および社会的弱者コミュニティの重点化などを通じて，後発開発途上国における気候変動関連の効果的な計画策定や管理の能力を向上するためのメカニズムを推進する．

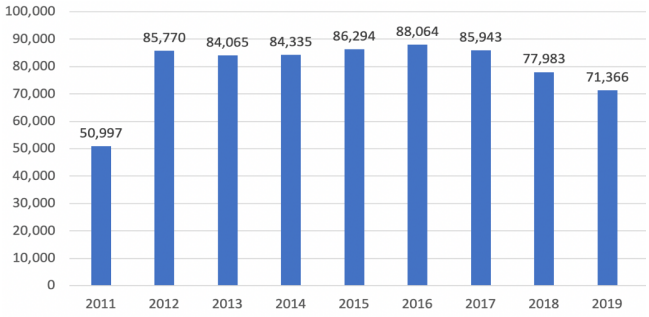

図 2.2-1 各年の新聞記事数

3. 提案手法

本章では日経コーパス中から SDGs のターゲットに該当する文を検索，時系列トレンドとして可視化する手法について述べる．

3.1 ベクトル生成手法：SWEM

本研究では，記事文と SDGs 目標のターゲットで構成されるキーワードを同じ特徴空間で表現できる必要がある．そこで文書の分散表現の生成手法として，Simple Word Embedding Based Models (SWEM) [7]という手法を採用する．この手法は単語の各次元の最大値や平均値を文書ベクトルとして採用する手法である．SWEM は優れた表現能力を持ちつつ学習パラメータを必要とせず，文章ベクトルを得るための計算コストも低い点で優れている．ここで本研究では，検索精度の結果から Word2Vec 自体は東北大学乾研究室の「日本語 Wikipedia エンティティベクトル」を使用する[10]．

3.2 ターゲット記事文の検索方法と時系列可視化

まず，新聞記事本文を MeCab [11] を用いて形態素解析を行う．この時，記事を文単位に分ける．次に，SWEM 法により文ベクトルを得る．次に，ターゲットについても，同様に文ベクトルを生成する．そして，時系列の記事文ベクトルとターゲットのベクトルの cos 類似度を取ることで，文中で SDGs に関するある話題がどの時期にどの程度語られているかについて，スコアリングした記事を時系列に並べることができる．本研究において，類似度のスコアが 0.9 以上のものを検索対象とした．

時系列トレンドは，検索対象となる記事文の算出したスコアを前後 4 週間ごとに集計・最大値で規格化することで話題の注目度がどの時期に高くどの時期に低かったのかという相対評価で表す．また，0.7 以上の話題に対して「強く語られている」と判別した．

4. 分析結果

4.1.1 検索評価

提案手法による，ターゲット毎に対する記事の検索能力について表 4.1-1 にまとめた．ここで，最終的に検索されるのはスコアの高い上位数アイテムなので，話題性の高い順に記事文を正しく並べ変えるタスクと捉えられる．そこで，ランキングの正しさを評価するため検索した 50 件の TopN の N=30 記事文において，① 実際にターゲットと記事文が対応している割合（適合率：Precision@N），② ターゲットと記事文が対応しているアイテムにおいて，レコメンドした N 個のアイテムが含まれる割合（再現率：Recall@N）そして，③ mAP（mean Average Precision）にて評価した．

その結果，適合率は目標 7 に関して 0.8 以上，目標 13 は 0.7，ランク指標も全体として 1 前後として SWEM による記事検索能力は信頼できる結果となった．

表 4.1-1 ターゲット毎の検索・推薦の評価結果

SDGs 目標 ターゲット	適合率 (Precision@N)	再現率 (Recall@N)	ランク指標 (mAP)
7.1	0.800	0.957	0.979
7.2	1.000	1.000	1.000
7.3	0.933	0.950	0.988
7.a	1.000	0.953	0.986
7.b	0.867	0.856	0.964
13.1	0.733	0.850	0.786
13.2	0.733	0.847	0.996
13.3	0.800	0.960	0.995
13.a	0.993	0.946	0.993
13.b	0.733	0.844	0.963

4.1.2 トレンド

図 4.1-1 SDGs に関する目標 7 (a)および 13(b)に紐づくターゲットの進行度とセクタを示す．両者ともに 2017 年以降に 0.7 以上の「強く語られている」シグナルが現れている．また，セクタの割合について資源・エネルギーが下降トレンドであるのに対して公的機関・大学の記事が上昇している．また，金融機関が対象期間にわたって一定の割合で言及されていることも特徴的である，

今回の分析結果を国際協力銀行（JBIC）の環境関連業務とそれを通じて把握される日本企業の事業動向から評価する．JBIC の環境関連の取り組みを計測するべく，環境関連のプロジェクトへの融資を分類した「環境」タグのついたプレスリリースの件数を，2010 年～2019 年について時系列で整理した（図 4.1-2）．これによると，2010 年には環境分野専門の融資プログラム「GREEN」が創設され，6 件が承諾されている．これを皮切りに，2014 年の 18 件まで基本的に増加傾向が継続するが，内談案件の一巡とともに 2015 年には 10 件，2016 年には 5 件へ減少した．その後，SDGs が発効した 2016 年を底に徐々に件数を回復してきており，2020 年は（コロナ禍でそもそも少ないリリース件数であったにもかかわらず）12 件と増加基調を強めている．

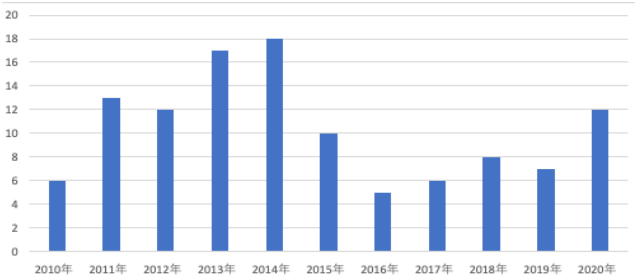

図 4. 1-2 JBIC の環境案件の件数（単位：件）

出所：JBIC プレスリリース

この間，海外向けの環境関連事業の政府支援は，途切れることなく実施されている．2011 年の東日本大震災を経て打ち出された「日本再興戦略 2013」を皮切りに，2017 年からは「未来投資戦略」と名称を変えていく．その中で，日本企業の海外戦略支援についても，経協インフラ戦略会議で「インフラシステム輸出戦略」が策定され，従来の機器売りからインフラをシステムとして売り込んでいこうという動きが強まっていき，さらに「インフラ」と「環境」が政策課題のキーワードとして緊密になっていった．この流れは現場を預かる JBIC 業務にも反映され，「地球環境保全業務」(2010 年 4 月)や「JBIC インフラ・投資促進ファシリティ」（2011 年 4 月）から，「質高インフラ環境成長ファシリティ（QI-ESG）」（2018 年 7 月）へと変化を遂げていく．このように，観測期間中，政策的支援が一定の後押しをしている中で，環境案件が S 字のトレンドを描いているという実態に鑑みると，今回の分析結果については大きな違和感はないものと言える．むしろ，現場の担当者が普段なんとなく感じているトレンドを可視化したという点で，意義のある分析であったと評価できよう．

5. 考察

SDGs に関する目標に紐づくターゲットの進行度の指標化，指標の時系列表示およびターゲットと最も類似する本文記事中の文の抽出ができた．一方で，扱う新聞記事データは，記者の取捨選択に伴っているため，本分析は日経新聞に限定されてしまう可能性は否定できない．今後は，国内外の新聞記事のみならず有価証券報告書などを用いて海外から見た SDGs/ESG の潮流と企業の課題，企業イメージを具体化する必要がある．

テキストマイニング手法は，人間が処理できないほどの大量な新聞記事データから，効率的にトピックを抽出することが可能であることが分かった．一方で，企業動向の詳細を人間が俯瞰し，整理することで定性的なインサイトを獲得する必要がある．

6. おわりに

自然言語処理の発展により，今まで対応できなかったスピード感で大量かつ粒度の細かい情報を処理することが出来るようになってきた．このように企業価値に占める無形資産，特にテキストデータの割合や重要性が近年高まってきており，それに伴い財務情報などからだけでは得られない情報をあらかじめセンシングする需要が高まっている．高度なアナリティクスを用いたパターン認識と機械学習は，従来人間が認識し得なかった要素をデータから意思決定に資する情報を的確に発見することが出来る．金融市場分析での機械と人間の当面の役割分担は，人間がまず関係のありそうなデータの範囲や目標を示し，そこから機械学習によるデータ解析を用いて有効なパターンの候補を機械に挙げてもらい，機械が提示した候補をどう評価して実際の投資に使うのかを人間が判断するというものになるであろう．さらに，過去データに無かったような新しいイベントや急激な変化が発生した場合の大局的な判断には，依然として人間の常識や直観による判断が求められると言える．

参考文献

[1] 和泉 潔，「ビッグデータと人工知能を用いたファイナンス研究の潮流」，『金融 研究』第 38 巻第 1 号，日本銀行金融研究所，2019 年，15～28 頁
[2] Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu, “Empirical Asset Pricing Via Machine Learning,” NBER Working Paper No. 25398, National Bureau of Economic Research, 2019
[3] United Nations, “Transforming our world: the 2030 Agenda for Sustainable Development”, 2015
[4] https://nkbb.nikkei.co.jp/alternative/article_data/
[5] World Economic Forum, “The Global Risks Report 2021 16th Edition”, 2021/19/1
[6] Kostovetsky, Leonard, and Jerold B. Warner, “Measuring Innovation and Product Differentiation: Evidence from Mutual Funds,” Journal of Finance, 2019
[7] Dinghan Shen, Guoyin Wang, Wenlin Wang, Martin Renqiang Min, Qinliang Su, Yizhe Zhang, Chunyuan Li, Ricardo Henao, and Lawrence Carin. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms. CoRR, Vol. abs/1805.09843, pp. 1–13, 2018.
[8] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. CoRR, Vol. abs/1405.4053, pp. 1–9, 2014.
[9] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Skip-thought vectors. CoRR, Vol. abs/1506.06726, pp. 1–9, 2015.
[10] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia 記事に対する拡張固有表現ラベルの多重付与. 言語処理学会第 22 回年次大会(NLP2016), March 2016.
[11] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto (2004) Appliying Conditional Random Fields to Japanese Morphological Analysis, EMNLP 2004

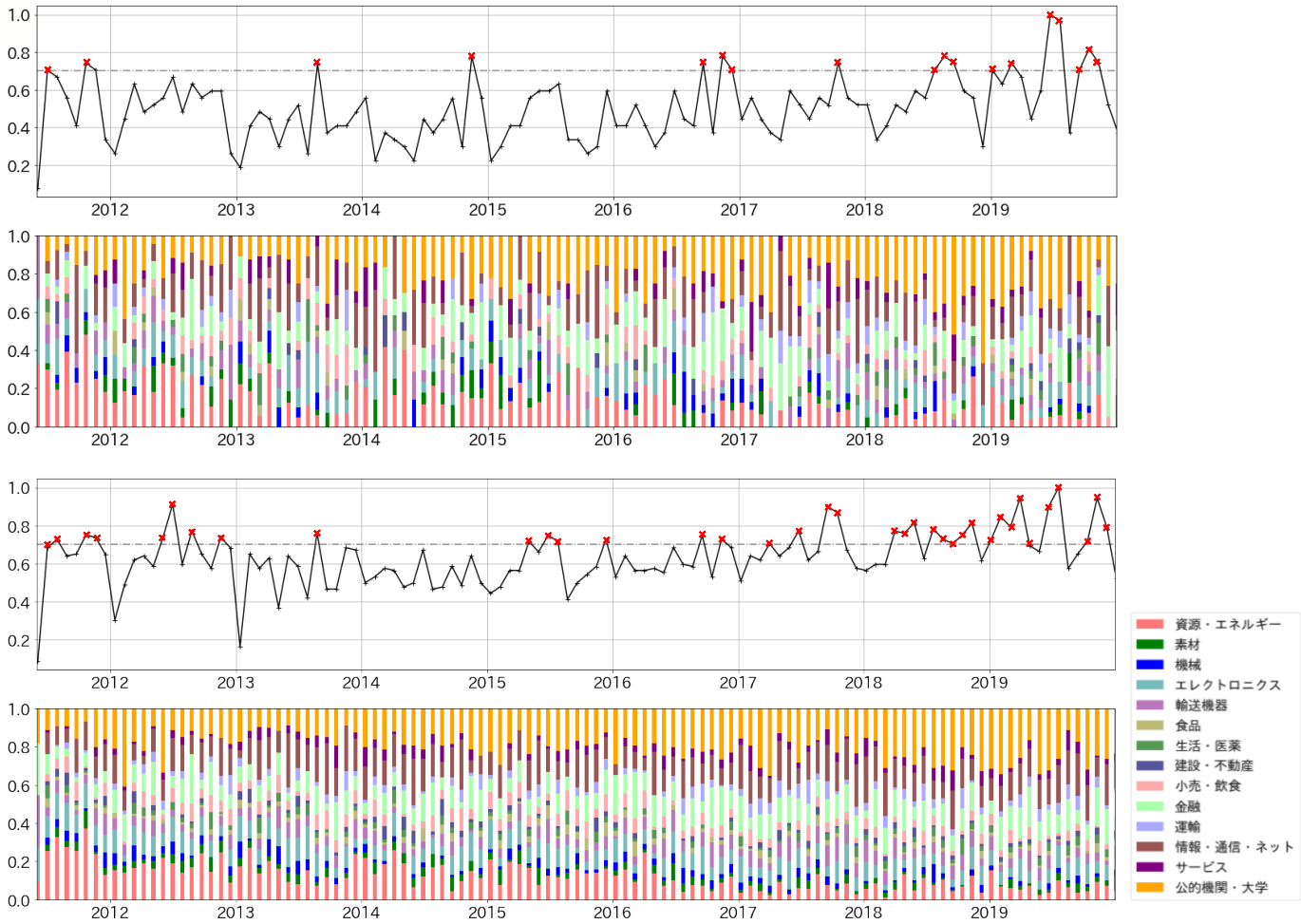

図 4.1-1 SDGs に関する目標 7 (a)および 13(b)に紐づくターゲットの進行度(上)とセクタの割合(下)

表 4.1-2 ターゲット 7.1 と話題との親和性

日付	記事文	類似度
2019/11/8	公共サービスを民間に開放する pfi はコスト抑制を通じて地域社会の持続可能性を高めるだけでなく，地銀の新たな事業機会にもなる．	0.937
2015/5/24	人々のニーズを満たすプル需要型ではなく，ニーズを創造するプッシュ供給型の資本主義は，社会的な成熟を回避させ，簡単単純早いことだけを消費基準とするキッザルトを必要とするのである．	0.934
2012/5/21	エネルギー取引の円滑化のため，インフラ発展の障害を除去し，よりクリーンで安全，受容可能な価格のエネルギーを追求する．	0.934
2012/9/9	増大する域内のエネルギー需要を安定的にまかなうため，資源の確保から効率利用までを網羅する．	0.933
2016/4/24	多様化するエネルギー源に対してコストや安全保障，持続可能性の点から包括的な役割を果たす組織になっていくことだろう．	0.932
2011/7/27	原子力，火力，再生可能エネルギーなどの発電コストを網羅的に再検証し，議論の土台となる客観的なデータを整える．	0.932
2011/6/18	安定的な電力を確保するためにも，エネルギー資源の調達先の多様化を急ぐ必要がある．	0.931
2017/7/30	宿命的な資源小国である日本のエネルギー政策には，大きく 1 輸入に頼らずエネルギーを確保する安全保障の視点 2 自然への負荷をできるだけ軽減する環境保護の視点 3 国民生活や産業競争力に資するため安価に供給するコストの視点という 3 つのポイントがある．	0.931
2014/4/2	富士通は低コストで森林状況を把握できるサービスの提供により，効率的な間伐を支援する．	0.929
2013/3/15	自由で競争的な電力制度のインフラとなるには，今後も飛躍的に取引を増やす必要がある．	0.929
2013/2/17	安定した供給を実現するには，市場の機能を活用することが基本的な解決策になると説く．	0.929

表 4.1-3 ターゲット 13.1 と話題との親和性

日付	記事文	類似度
2011/6/15	原発のあり方や再生エネルギーの活用を含む全体像について，コスト，技術進歩，環境への影響など総合的な観点から考える必要がある．	0.924
2019/5/5	エネルギー問題は安全性や環境への対応，経済性，安定供給など様々な要素を考える必要がある．	0.924
2014/10/24	地震や火山の影響を避けるための科学的な分析に加え，人口密度や輸送のしやすさ，周辺環境の保護などを考慮して適地を絞り込む．	0.922
2016/2/9	近年は水や土壌の汚染対策にとどまらず，生物多様性の保全など各国の安全基準が厳格化．	0.921
2013/1/17	エネルギー需給の安定や安全保障，気候変動対策など中長期的なエネルギー戦略の一環として，省エネの重要性が改めて認識されるべき段階に来ている．	0.921
2015/5/26	安全性についてはユーザーの理解促進や許容可能なリスクを洗い出すこと，消防関係者への講習徹底など，規制以外の環境の整備も求められよう．	0.921
2013/3/5	こうした教訓を踏まえ，産業構造やエネルギーだけでなく，情報通信，金融，交通物流など幅広い分野にまたがるインフラの災害リスクを検証する．	0.920
2017/10/31	温暖化による災害対策としては，風水害のリスクや農業が受ける影響を予測する方法を開発する．	0.920
2016/3/7	また再生可能エネルギーや原子力を現状より拡大する必要があるが，コストや社会的な受容性の面でともに弱点を抱える．	0.920
2012/9/12	原発内外の放射性物質による汚染や住民の被曝ひばくなどのリスクに対し安全確保を重視する方針も示した．	0.920