

日本の力を、世界のために。

Supporting Your Global Challenges



ニュース記事を用いた企業動向分析の試み

Japan Bank for International Cooperation
Strategic Research Department

2021年11月

(株) 国際協力銀行 (Japan Bank for International Cooperation : JBIC)

- 本店所在地：東京都千代田区大手町1丁目4番1号
- 資本金：1兆9,638億円（日本政府が全株保有）
- 職員数：657名

企画部門> 調査部> 第1ユニット

● 「海外事業展開調査」

日本の製造業500～600社を対象に「海外事業展開調査（アンケート調査）」を毎年行い、今後の海外事業展開に対する姿勢を分析。調査結果については、国内外に広く報告している。各国大使館や海外駐在員事務所との連携、国内外でのセミナーも開催。



最新の報告書はこちら 

沿革



株式会社国際協力銀行 (本店)

〒100-8144 東京都千代田区大手町1丁目4番1号
TEL: 03-5218-3100
FAX: 03-5218-3955

ウェブサイト <https://www.jbic.go.jp/>

日本の力を、世界のために。

Supporting Your Global Challenges

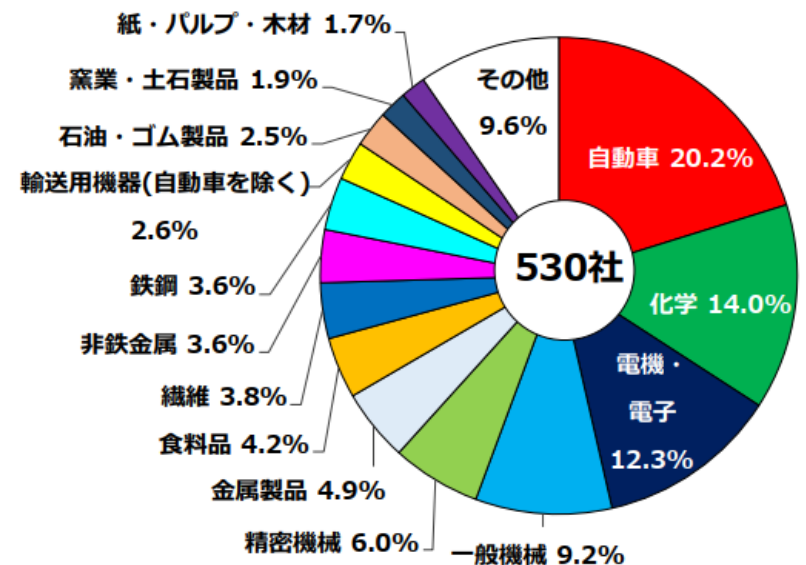


1. 動機と手法

- ✓ 1989年から実施している「海外事業展開調査」の対象は製造業で、調査対象に非製造業も含まれないか、行内外から強い希望あり。
- ✓ ただし、調査手法（各企業へのアンケート形式）を援用するかどうかが課題。非製造業企業を対象にアンケートを実施した場合、各企業へ負担をかける恐れ。また、そもそもアンケート形式での調査は、テレワークの進捗、コンプライアンス意識の高まり等で年々難しくなっている。



上記の観点から、①企業に回答の負担をかけない形で、②製造業・サービス業を網羅する、新しい企業動向分析（マーケティング）手法を開発する必要性 → 今回の取組みの発想



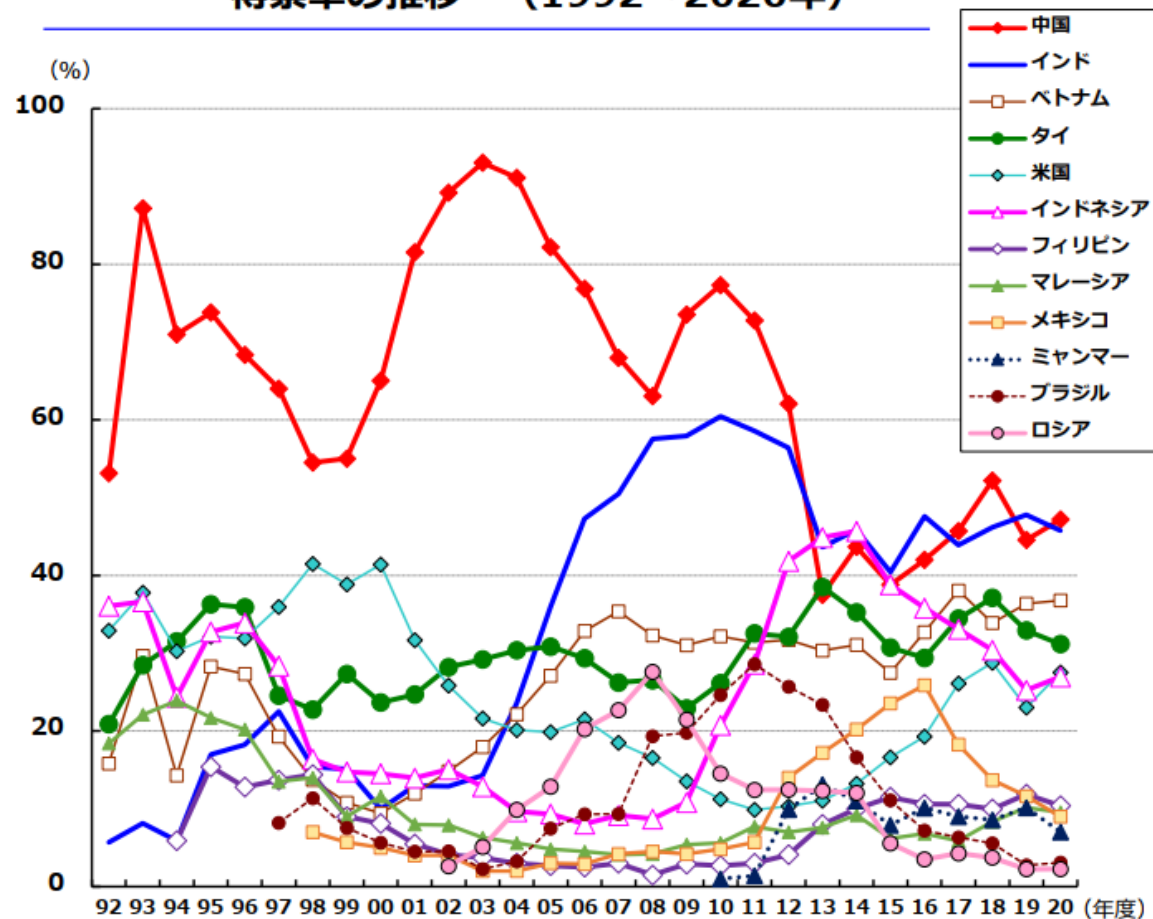
- アンケート調査の重要なトピックが「有望国ランキング」。日本企業の今後の海外進出を占うもので、各有望先国を自由記入形式で回答してもらい、集計したもの。この調査を、過去のテキストデータを用いて、何らかの形で代替することがプロジェクトの基本的なゴール（の一つ）。

中期的な有望事業展開先国・地域（今後3年程度）

※得票率 (%) = 当該国・地域の得票数 / 本設問への回答社数

順位		国・地域名 (計)	回答社数 (社)		得票率 (%)		
2020	← 2019		2020	2019	2020	2019	
			356	404	2020	2019	
1	↑	2	中国	168	180	47.2	44.6
2	↓	1	インド	163	193	45.8	47.8
3	-	3	ベトナム	131	147	36.8	36.4
4	-	4	タイ	111	133	31.2	32.9
5	↑	6	米国	98	93	27.5	23.0
6	↓	5	インドネシア	96	102	27.0	25.2
7	-	7	フィリピン	37	48	10.4	11.9
8	↑	9	マレーシア	34	41	9.6	10.1
9	↓	8	メキシコ	32	47	9.0	11.6
10	↓	9	ミャンマー	25	41	7.0	10.1
11	↑	14	ドイツ	20	14	5.6	3.5
12	↓	11	台湾	18	18	5.1	4.5
13	↑	21	バングラデシュ	16	7	4.5	1.7
14	↑	15	オーストラリア	14	13	3.9	3.2
15	↓	12	韓国	12	15	3.4	3.7
16	↓	12	シンガポール	11	15	3.1	3.7
16	↑	17	ブラジル	11	11	3.1	2.7
18	↑	26	英国	9	4	2.5	1.0
19	↓	18	ロシア	8	9	2.2	2.2
20	-	20	トルコ	7	8	2.0	2.0

得票率の推移（1992～2020年）



JBIC側

- ・ 調査責任者
 - ・ 担当者×2
-
- ✓ テキストマイニングの実績なし
 - ✓ 主に企業財務分析、マクロ経済分析
 - ✓ 融資向けの企業審査・ソブリン審査

パートナー側

- ・ マネージャ×1
 - ・ データアナリスト×3
-
- ✓ テキストマイニング分析の経験豊富
 - ✓ TMSの利用に習熟
 - ✓ Pythonなどのプログラミングにも通暁

分析環境／期間

JBIC執務室内に計算機（ノートPC）を用意、調査期間は2020年10月～12月

- 結局、マシンスペックが足らず、パートナー側の環境を利用。
- 導入したTMSはスタンドアロン版。クライアントサーバ版にすればよかったかもしれない

- **データ**：本調査の分析対象となるテキストデータの要件は、分析手法やその有効性を確認するべく、より簡便で結果まで最短でたどり着くためのデータ、という観点で選択。以下の条件を満たすデータとして、日経オルタナティブデータ（新聞記事をテキスト化し、タグ付けしたデータ）が浮上。
 - ✓ 日本語である
 - ✓ 文章の表記ゆれが少なく、用いる単語の質が安定している
 - ✓ 情報の信用性・即時性（＝テキストデータのタイムスタンプが精確）が高い
 - ✓ 日本企業に関する情報量が多い
 - ✓ 時系列分類が済んでいる
 - ✓ 会社名、業種、キーワード（「設備投資」「研究開発」）等、多様なタグが付与されている
 - ✓ 長期スパンで入手可能

- **解析ツール**：調査担当者がいずれもプログラミングについての知識が無いため、初心者にも使いやすいプログラムを検討。以下の観点から、NTTデータ数理システムのTMSを選定。
 - ✓ 日本語の特徴をふまえたテキスト解析に適している（係り受け解析など）
 - ✓ 日本語で操作可能
 - ✓ 導入の研修が充実している
 - ✓ 導入の初期費用が予算内
 - ✓ 結果をグラフィカルに表示できる（ことばネットワークなど）

□ **結果**：様々な試行を重ねる中で、結局以下のような手法をとった。

- ✓ **検証した分析手法**：4つの手法を検証。具体的には、①**注目度分析**、②**トレンド分析**、③**ポジネガ分析**、④**ベクトル分析**（Word2Vec）、をそれぞれ試行。
- ✓ **使用したデータの量**：日経新聞の過去10年分の記事（※朝刊のみ。社会面・スポーツ面を除外した）。実際に使用したのは、上記手法のうち①～③については**1万4,000記事（約800万字）**、④については**全74万記事（約5～6億字）**が分析対象となった。
- ✓ **データクレンジングの工数**：上記手法の①～③については、対象データの選定を実施。具体的には、日経データの「業種タグ」や海外事業に関係するタグを活用し、また記事内容についてもクレンジング（国名・企業名の名寄せ、ノイズとなるスペースやカッコ等の削除）を実施。最も工数を必要としPythonでの処理を要したが、あらかじめタグ付けされていたぶん、工数を圧縮。
- ✓ **分析ツール**：用いたプログラムなどは、Text Mining Studioに加え、Python、Word2Vec（ベクトル分析）、ASARI（ポジネガ分析）など。



詳細は次頁以降

日本の力を、世界のために。

Supporting Your Global Challenges



2. 分析結果



- ◆ **国の顔ぶれは納得感あり**：企業活動と相関性の高い記事を抽出する処理を施し、**記事への登場回数をランキング形式で集計**。具体的には、対象とする記事について、71の共起単語を事前に指定する処理を行い、精緻化したもの。この順位そのものを非製造業の有望国ランキングとして使うことはできないが、出てくる国の顔ぶれはアンケート調査に類似している。記事データを用いた機械的な処理でも、一定の納得感のある国名の集合は得られると判明。
- ◆ **活用の方向性**：機械的に迅速な処理が可能で、変化の見える化や注目すべき時期などの「気づき」を得るところまでは効率的に作業できる。そこから先のインプリケーションの抽出にあたっては、結局人の目を通じなければならぬため、できるだけ人の目を介さなくても良くなるよう、機械的に行う前処理の段階から更なる精度向上の余地あり。

日本の非製造業が注目している国はどこか

順位	2011年	2012年	2013年	2014年	2015年	2016年	2017年	2018年	2019年	2020年
1位	中国	中国	米国	米国	米国	米国	米国	米国	米国	米国
2位	米国	米国	中国	中国	中国	中国	中国	中国	中国	中国
3位	タイ	タイ	タイ	ベトナム	タイ	インドネシア	ベトナム	タイ	韓国	タイ
4位	ベトナム	インドネシア	インドネシア	タイ	ベトナム	ベトナム	インド	ベトナム	インド	インド
5位	韓国	ベトナム	ベトナム	インドネシア	インド	台湾	インドネシア	インド	ベトナム	韓国
6位	インド	インド	インドネシア	シンガポール	シンガポール	タイ	タイ	英国	インドネシア	シンガポール
7位	インドネシア	韓国	台湾	台湾	インドネシア	シンガポール	シンガポール	オーストラリア	英国	ベトナム
8位	英国	台湾	ミャンマー	インド	韓国	インド	韓国	インドネシア	シンガポール	英国
9位	シンガポール	シンガポール	マレーシア	マレーシア	香港	韓国	台湾	シンガポール	タイ	ドイツ
10位	台湾	英国	シンガポール	ミャンマー	ミャンマー	マレーシア	英国	韓国	ドイツ	フランス

1. この手法による調査結果

- 順位そのものについてはアンケート調査との相関性は高くなく、海外投資アンケートに代わるものではない。ただし、出てくる国の顔ぶれはアンケート調査と似ていることから、記事データを用いた機械的な処理でも、一定の納得感のある国名の集合は得られるという事が分かった。
- ランキングのインプリについては、実際の投資額と注目度分析との相関性を高めるなどの工夫の余地あり。

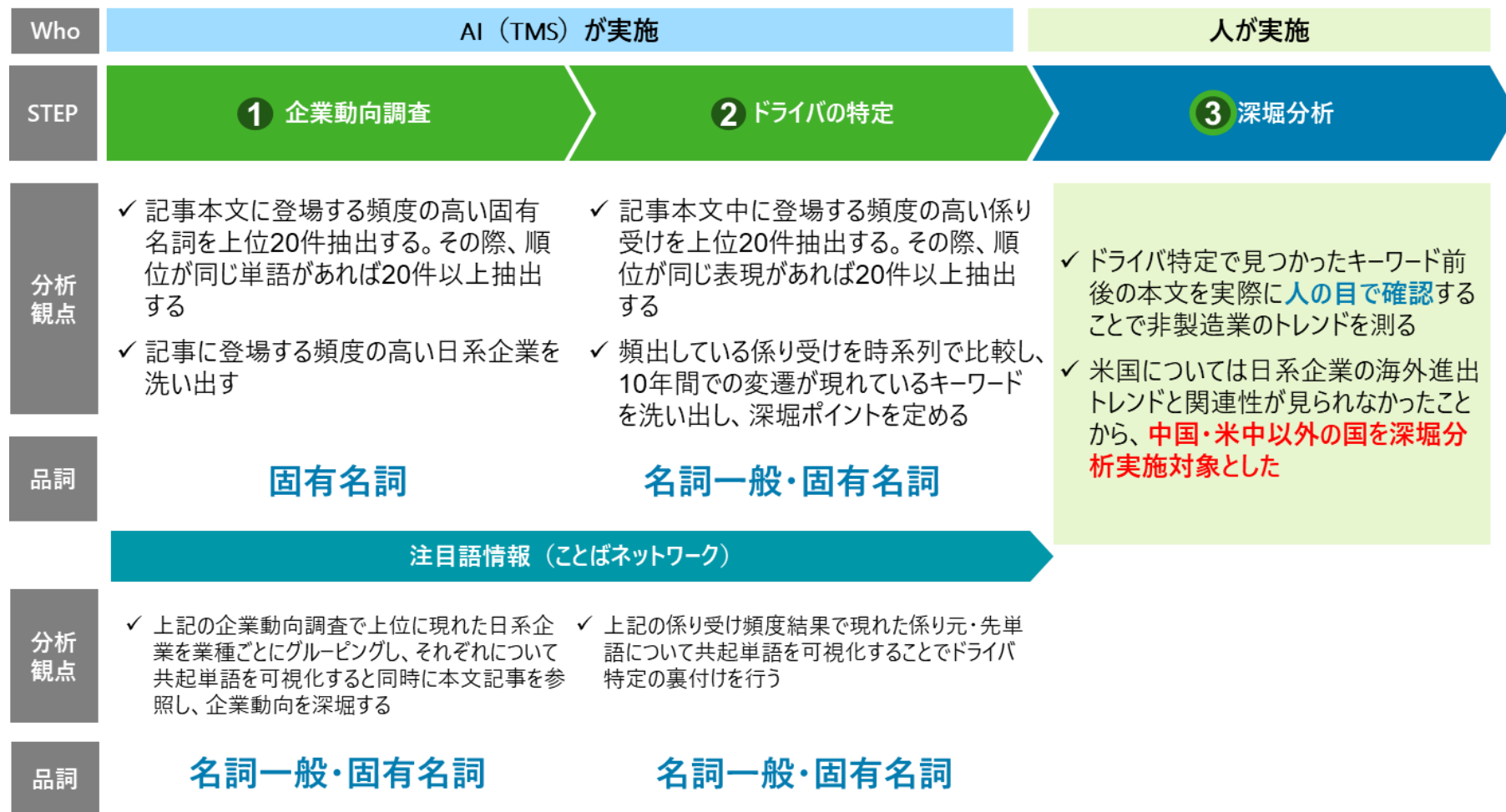
2. この手法の特長

- 人間ではとても処理できない膨大な量の記事データから、注目される国や変化の時期について、機械的に、かつある程度の納得感をもって見える化が可能。
- また、ランクインしている国の実際の記事に目を通して内容を確認することもできるため、注目国の順位が変化した背景や理由について、迅速に確認することができる。

3. この手法の課題

- 新聞で記事化されない情報は出てこない。また、新聞で記事化されやすい国に偏りがち（記事数：アメリカ3,914件、中国3,445件、タイ：895件、他）。
- 変化の見える化や注目すべき時期などの「気づき」を得るところまでは効率的に作業できるが、そこから先のインプリケーションの抽出にあたっては、結局人の目を通じなければならない。
- 更なる精度向上の余地あり。様々な設定やデータ抽出の工夫の他、SOE（System of Engagement）という技術（所謂ウェブサービスの「おすすめ機能」）を用いて、注目すべき国の抽出を（日常的に）AIに提案させるといった改善が検討余地あり。
- また、技術的には予測モデルを作成することは可能。

□ **トレンド分析**：時系列で国別及び業種別のトレンドの推移を大量の記事データから見える化する手法。具体的には、①**個別の企業名**と、②**トレンドを動かすドライバ（共起単語）**、の双方について、TMSを用いて注目すべき単語情報を特定。その後、③**人の目を通して記事を確認しつつ深堀分析**を実施。

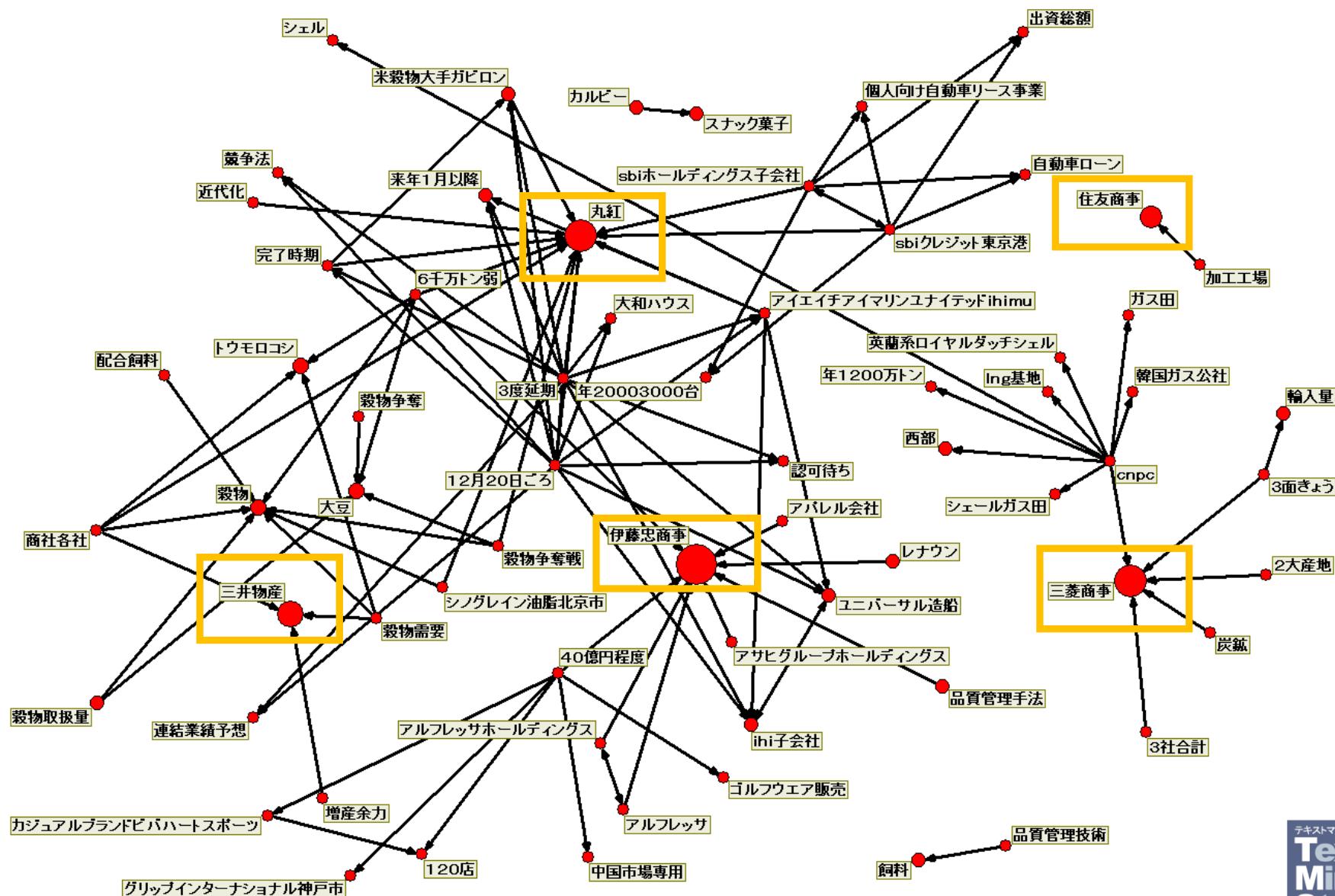


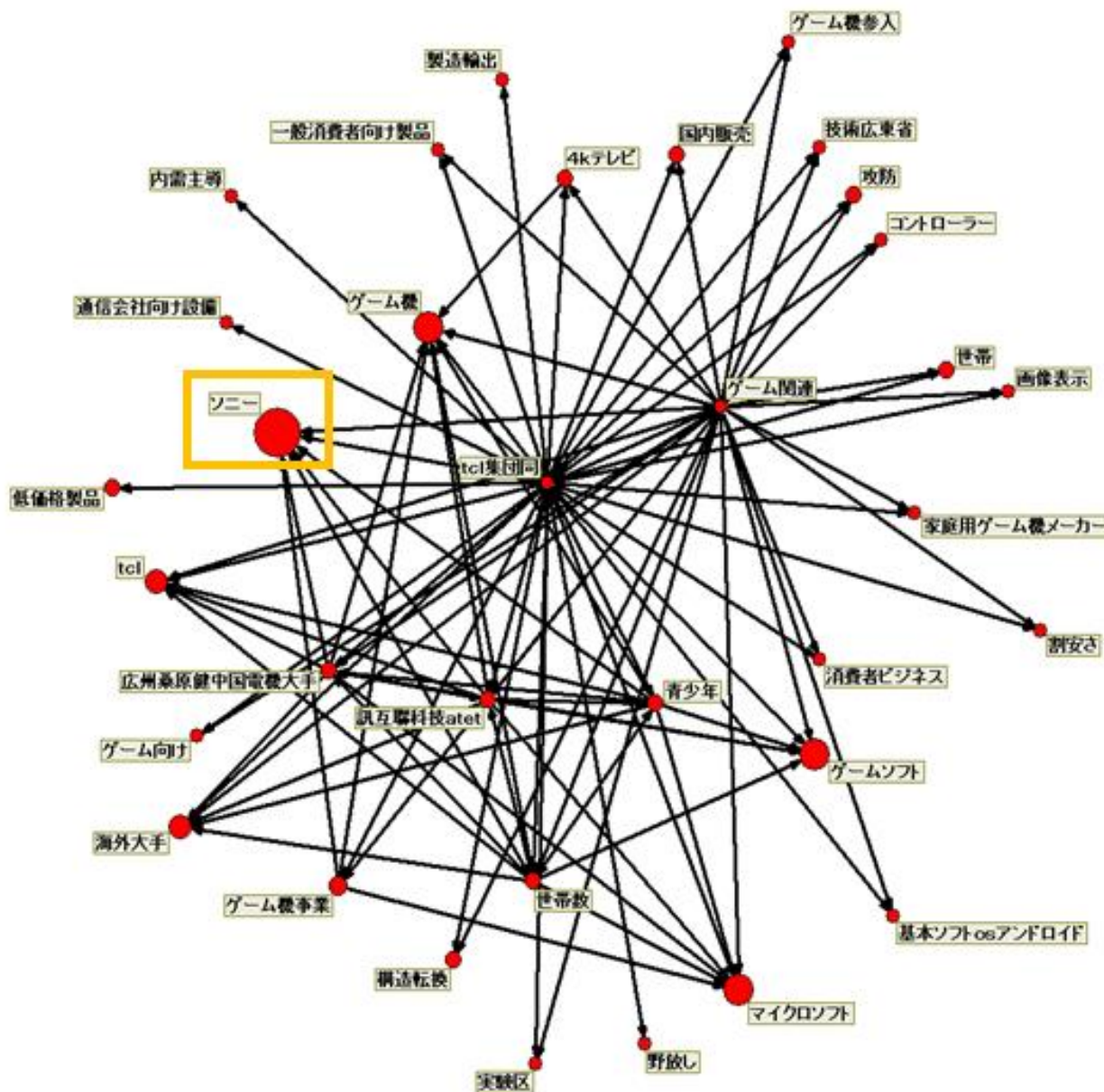
2. 分析結果 ②トレンド分析 企業やその動き方の特徴を抽出

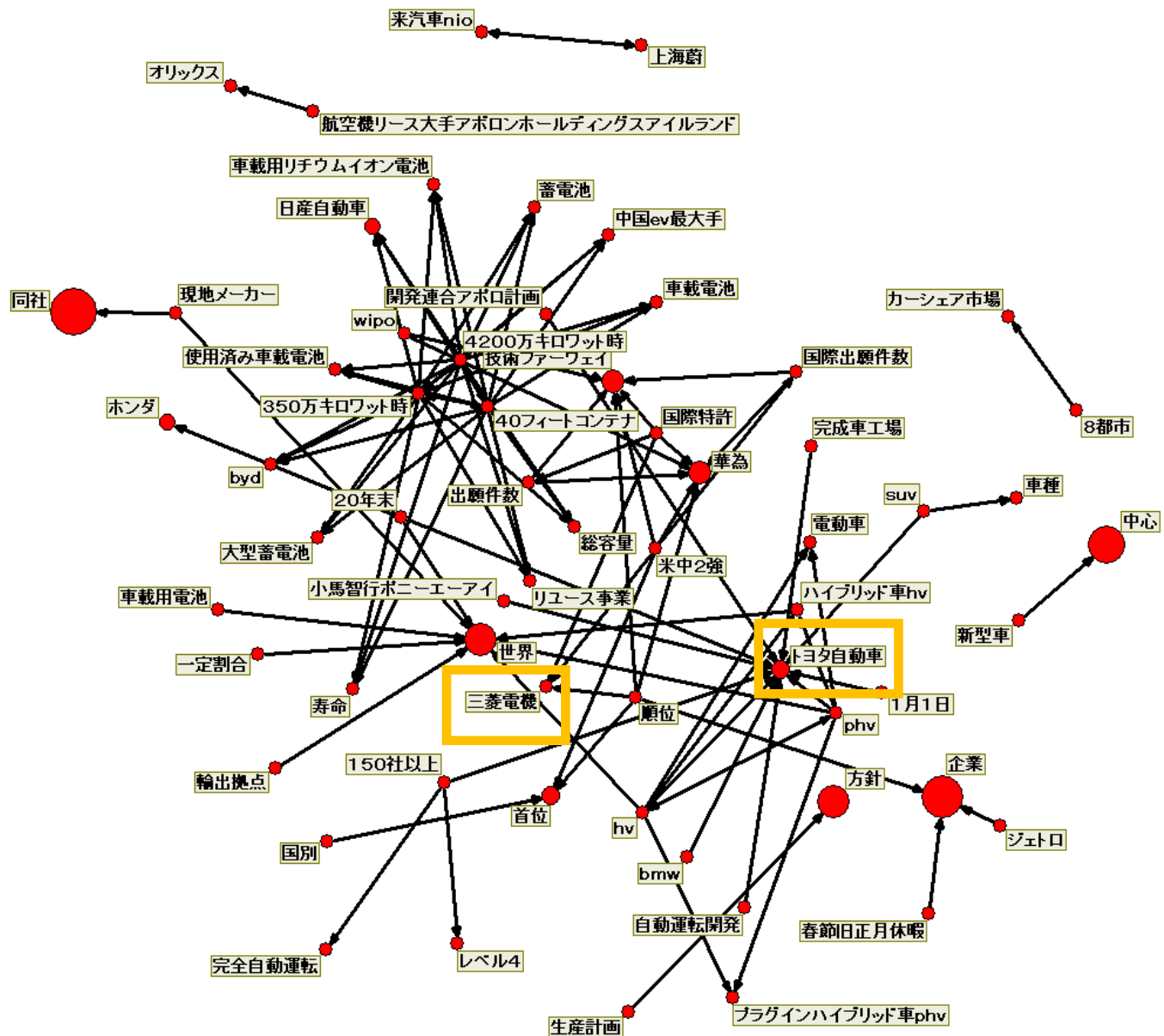
◆ **企業名とトレンドの抽出が可能**：企業活動のトレンドについて、その**ドライバ（単語）を可視化**することで、分析者に気付きを与える手法。例えば今回、中国を対象とした分析において、伊藤忠やトヨタが活発に動いており、また「富裕層から若者へ」「都市部から地方へ」というトレンドが確認できた。いずれも概ね納得感のある結果が抽出できている。

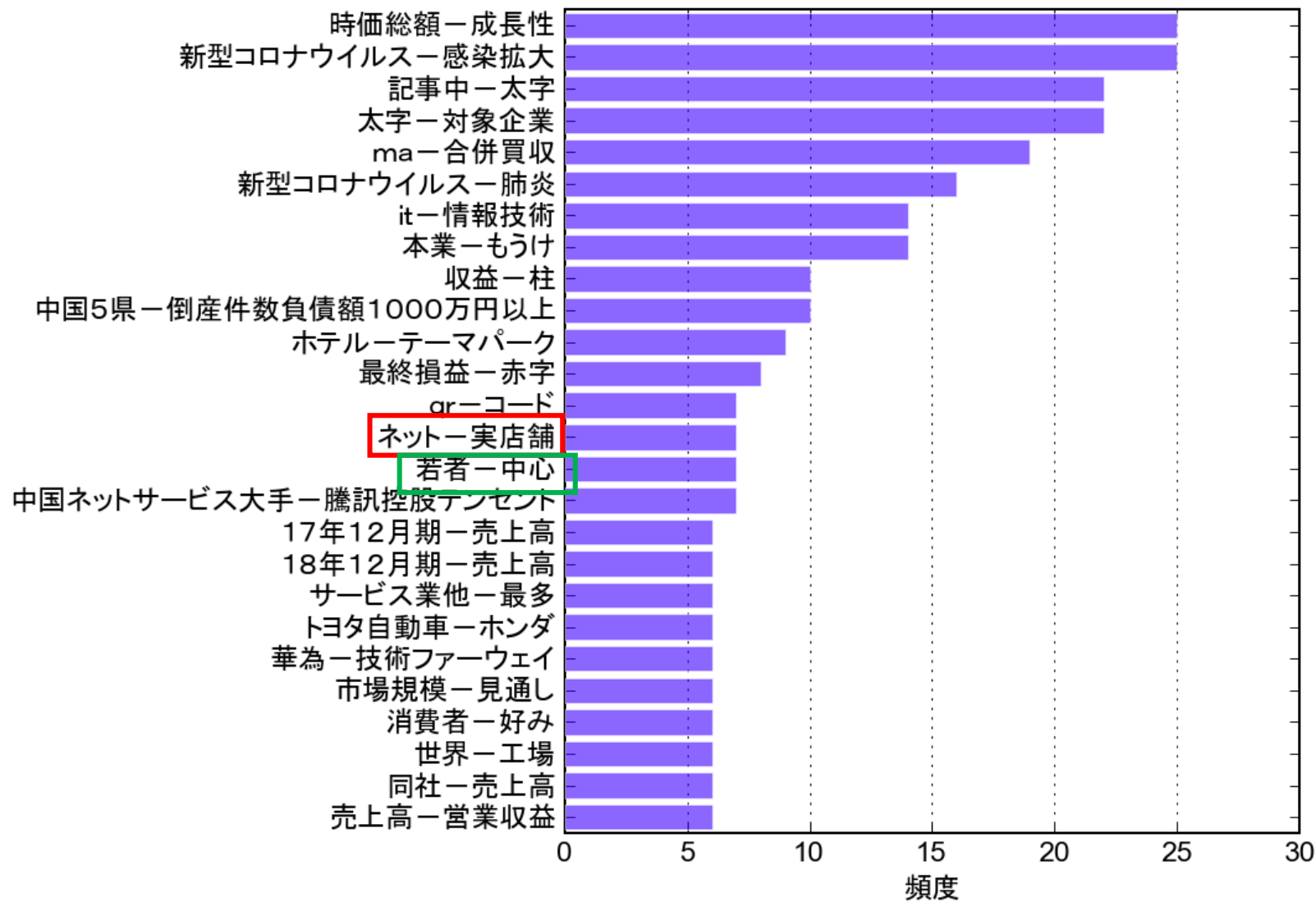
◆ **活用の方向性**：今回、ある程度の納得感のある結果が引き出せたことは、この手法（使用したデータマートとツール）の潜在能力、すなわち多様なテーマに対してこの手法での最新トレンド分析が可能であることを示唆している。例えば経営計画に掲げたテーマと実際の企業動向等との相関や先行性を確認することにも利用可能と考えられる。

2011年～2013年		2014年～2016年		2017年～2020年		2011年～2013年		2014年～2016年		2017年～2020年			
順位	単語	頻度	単語	頻度	単語	頻度	順位	係り元単語	係り先単語	頻度	係り元単語	係り先単語	頻度
1	伊藤忠商事	45	伊藤忠商事	51	トヨタ自動車	39	1	ma	合併買収	13	ma	合併買収	26
2	丸紅	32	ユニクロ	19	伊藤忠商事	33	2	取引先	企業	13	時価総額	成長性	16
3	三菱商事	32	三井物産	15	グーグル	24	3	収益	柱	9	若者	中心	10
4	パナソニック	23	ソニー	14	韓国サムスン電子	23	4	都市部	中心	9	化粧品	日用品	7
5	三井物産	22	トヨタ自動車	11	ホンダ	21	5	oem相手先	ブランド	8	現地企業	合併	7
6	トヨタ自動車	19	ヤフー	11	日本経済新聞社	19	6	現地	旅行会社	7	事業	柱	7
7	日産自動車	18	経済産業省	11	パナソニック	18	7	新興国	中心	7	品質	高さ	7
8	ユニクロ	17	住友商事	11	経済産業省	14	8	製造業	中心	7	oem相手先	ブランド	6
9	住友商事	16	韓国サムスン電子	10	ソニー	13	9	富裕層	中心	7	現地	旅行会社	6
10	ヤマダ電機	15	三菱商事	10	ファーストリテイリング	12	10	格安航空会社	春秋航空	6	現地企業	合併会社	6
11	ソニー	14	グーグル	8	ユニクロ	12	11	企業	ニーズ	6	人気	商品	6
12	韓国サムスン電子	14	丸紅	8	ロイター	12	12	現地	消費者	6	中国インターネット検索最大手	百度バイドゥ	6
							13	事業	柱	6	不動産大手	大連万達集団ワンダグループ	6
							14	成長戦略	柱	6	パソコン	スマホ	5
							15	大都市	中心	6	家庭用ゲーム機	製造販売	5
							16	日本製品	不買運動	6	華為	技術ファーウェイ	5
							17				収益	柱	5
							18				地方都市	中心	5
											消費者	好み	6









1. この手法による調査結果

- 中国を対象とした分析では、「富裕層から若者へ」「都市部から地方へ」というキーワードが非製造業のトレンドを形成しているなど、納得感のある結果が抽出できた。
- 深堀分析の枠組み次第では、様々な業種の動向を追うことができ、汎用性が高い。
- 言葉ネットワークでは、個別企業の動向を特徴づけるような、トレンドが可視化された。

2. この手法の特長

- トレンドのドライバを特定し可視化することで、人に気付きを与え、分析をより効率的に、かつエビデンス（元記事）があることでより説得的な内容にすることができる。
- 結果を見る限り驚きはないが、逆にこうした納得感のある結果が引き出せたことは、この手法（使用したデータマートとツール）の潜在能力、すなわち多様なテーマ（例えば中計で掲げたテーマに関する実際の企業動向等）に対してこの手法での最新トレンド分析が可能ではないか、ということを示唆している。

3. この手法の課題

- 新聞で記事化されない情報は出てこない。
- 頻出単語が抽出されたとしても、単語の使われ方・文脈（記事の趣旨）までは抽出できないので、トレンドやドライバを特定した後は、結局のところ人の目を要する深堀分析が必要になってくる。

ポジネガ分析：特定の国に関する記事が、ポジティブ/ニュートラル/ネガティブの3カテゴリのどれかを分類する。感情分析プログラム（ASARI）によって数値化し、その時系列の推移を国軸・業種軸で見、その国のイメージの変化を傾向として捉えることを試みる。

<参考：分析手順>

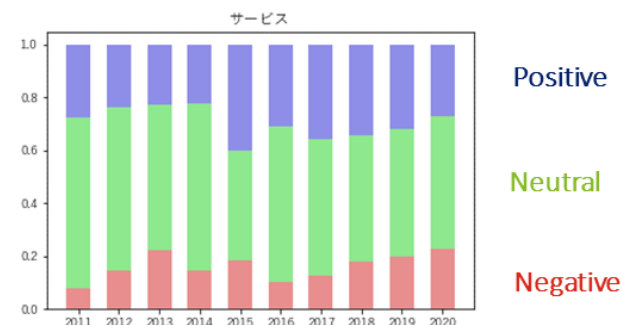
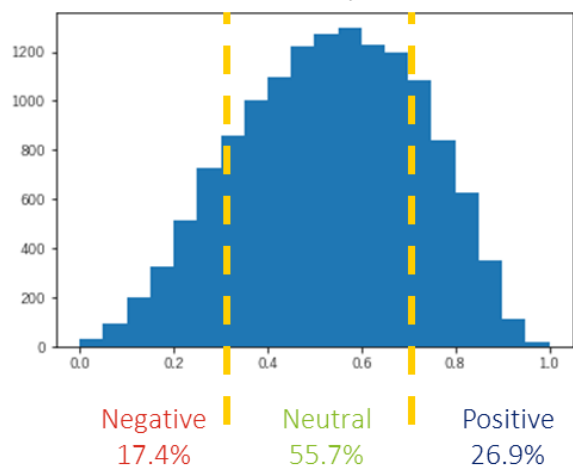
記事を分類（Positive / neutral / negative）

- 機械学習手法を用いて、記事がpositiveかnegativeかの2クラスに分類し、その確信度を算出
- Positiveの確信度とnegativeの確信度は合計で1になるようになっている
- 今回の分析では、positiveの確信度をpositive度合いを表すスコアとして扱った
- スコアの高い方からpositive（スコアが2/3以上）、neutral（1/3以上2/3未満）、negative（1/3未満）という3クラスに分類を行った

時系列で各カテゴリの推移の可視化

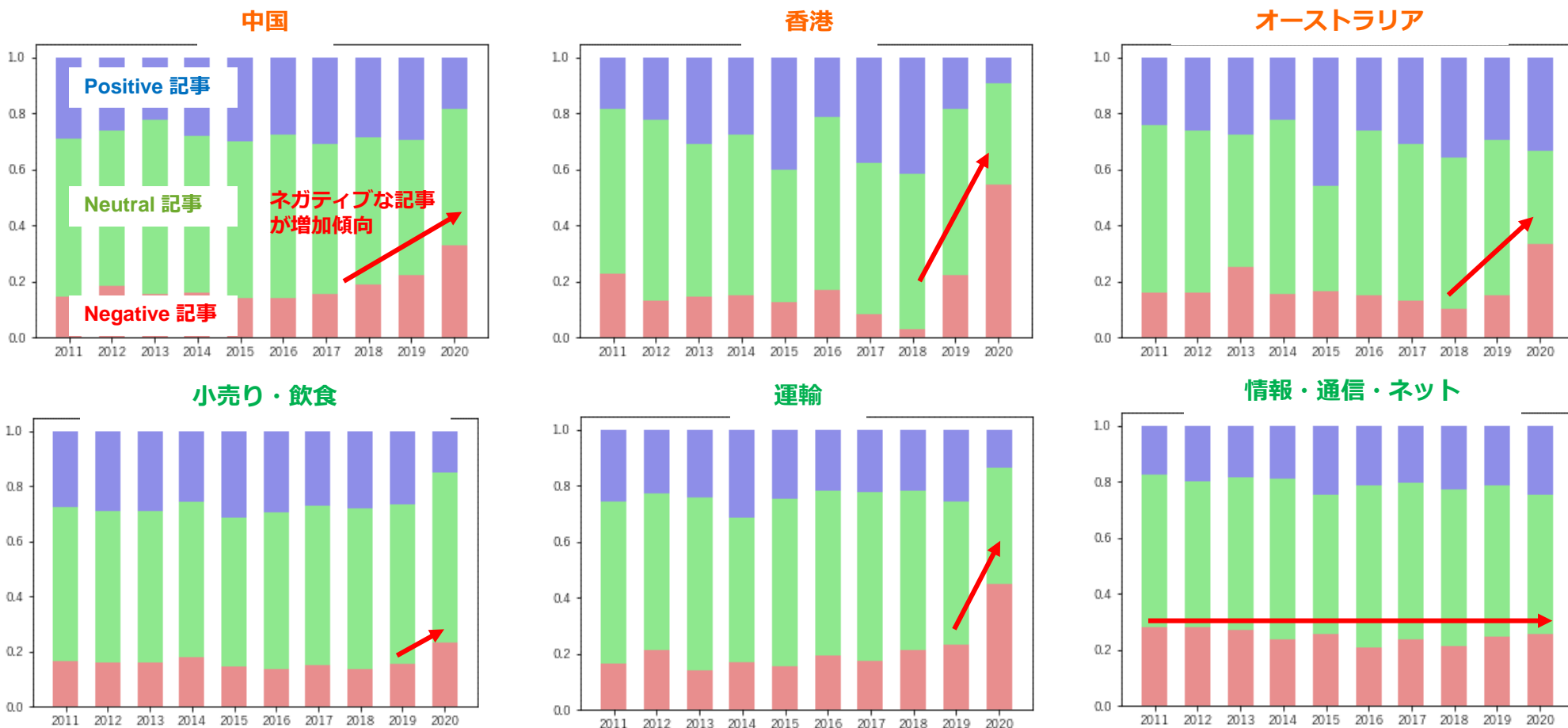
- 左記の3クラス分類の推移を年毎に集計し、時系列で可視化
- 年毎に記事数のばらつきが存在するため、年内の記事における各クラスの割合に変換して可視化
- 国軸については、50件未満の国は除外

データマート全体に対するpositiveスコアの分布



2. 分析結果 ③ポジネガ分析 リアルタイムな変化を捉える → アラート機能向け JBIC

◆ **簡便なトレンド把握／リアルタイムな分析に有効**：国別では、中国、香港、オーストラリアなどでネガティブ割合が増加していることを捉えており、昨今の米中対立の影響の大きさを示している。また業種別では、小売りや運輸でコロナ禍の影響を捉えている様子がわかる。以上から、大まかな傾向や、イベント発生時の影響をリアルタイムに把握したい場合に、期待できる手法だと判明。



1. この手法による調査結果

- 国別・業種別とも、大まかな傾向をつかむのに適した手法であることがわかった。
- 直近のコロナウイルスの影響について、リアルタイムで把握することができている。
- ポジティブ/ニュートラル/ネガティブの分類の精度が高く、信頼性は高いことが分かった。

2. この手法の特長

- 特定の国や業界のイメージ（=日本での報道のされ方）を大まかに把握し、時系列で比較することでイメージのトレンドや転換点を把握することができる。このため、各国の事業環境に対する日本の経済界の印象やイメージのトレンド転換を素早くつかむことが可能。
- ポジネガという人間でも判断が難しいタスクを一律かつ効率的に処理することで、横比較が可能となる。このため、今後の活用案としてアラート機能（=リスク管理機能）を持たせ、足元の動きを随時追うような機能を構築することも考えられる。その場合はデータマートの頻度を高めるなどの工夫が必要。

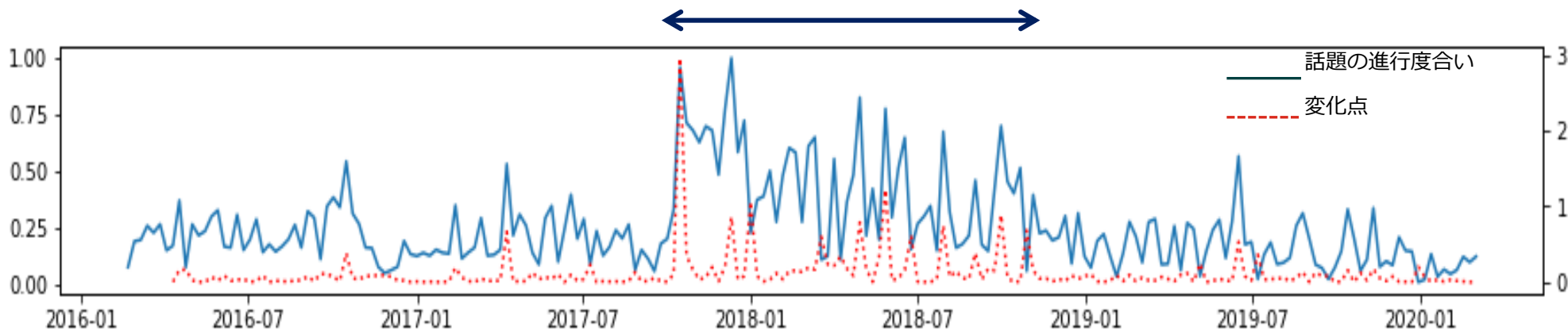
3. この手法の課題

- 特定の媒体（日経新聞）のデータなので、ポジネガ判定が政治的な背景に偏りがち。

- ◆ **SDGsの注目度合いを可視化**：様々なプログラムが開発されているが、今回は深層学習をベースにした Word2vec というプログラムを用い、10年分の全記事をベクトルに転換、SDGsに関する17の目標で語られているテキスト情報との類似性を検証し、時系列で整理することで、SDGsの各ゴールの“バズり具合”のグラフ化を試みたもの。
- ◆ **手順**：形態素解析には MeCab を使用。記事コーパスを期間別（四半期、年度毎）に用意し、それぞれの期間別に学習した Word2Vec を作成。単語ではなく文書が持つトピックをベクトル表現として扱い、類似度を算出することで適合率および再現率の精度を向上させた。

トレンドをけん引している要素（ドライバー）を洗い出す

トレンドを牽引しているドライバのシグナル

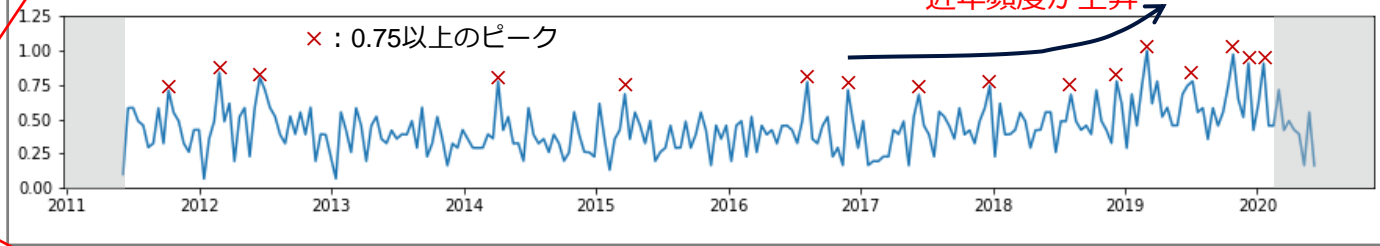


(1) 目標7：エネルギーをみんなに そしてクリーンに

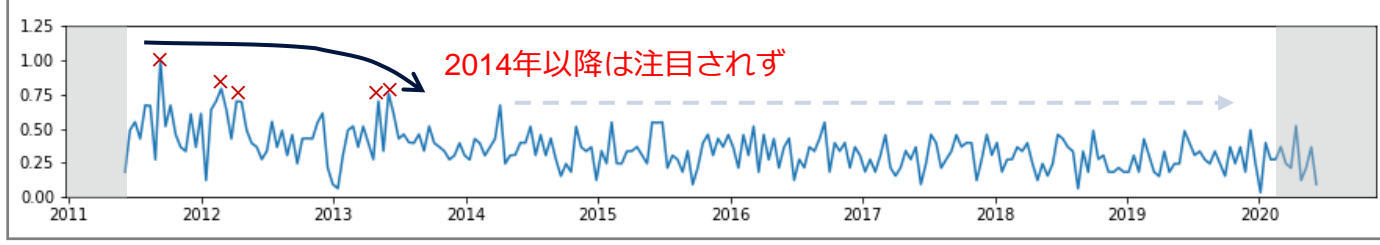


目標7：
世界中のすべての人々が、クリーンで安価なエネルギーにいつでもアクセスできることが目標。2030年までに電力を全世界に普及させるためには、化石燃料への依存から脱し、太陽光や風力、地熱など代替的なエネルギー源に移行することが重要。

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.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、内陸開発途上国のすべ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”



この文章を丸ごとプログラムに入れてベクトル化し、日々の記事データとの相関度合いを時系列で可視化した。

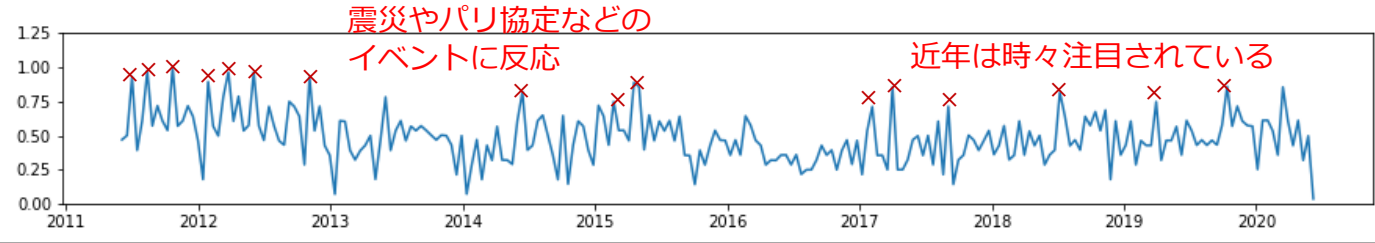
※グラフは2週間の移動平均なので、「×」印は単に1記事ではなく、その前後で注目度が平均的に高まっていることを示している。

(2) 目標 13：気候変動に具体的な対策を



目標13：「気候変動に具体的な対策を」は持続可能な生産消費形態を確保することが目的である。2030年までの天然資源の持続可能な管理と効率的利用、食糧廃棄や廃棄物の削減、化学物質などの放出の低減が謳われている

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”



13.3 “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”



1. この手法による調査結果

- 目標7については、「クリーンで安価なエネルギーへのアクセス状況」の観点で、企業の意識が近年急速に高まっていることが判明した。他方、「インフラの支援」に関しては以前は少し話題になったものの、近年はあまり盛り上がっていないという結果。目標13についても、SDGs以前ではイベントドリブンであったが、2016年頃から近年にかけて徐々に話題が上る様子が把握できた。いずれも、概ね違和感はない結果となっている。
- なお、その他の目標についても、同様の分析を実施済。カテゴリによっては、「SDGs」という単語の出現以前から取り組みがあったことが分かった。

2. この手法の特長

- 約74万件という膨大な記事データに関して、設定した話題との親和性を時系列で効率的にピックアップ。トレンドを把握した上で、具体事例を収集する上でも非常に効率的なツールとなり得る。
- 最近注目され始めた単語やトレンドについて、事後的にその源流を遡ることができる。
- トrend分析では3段階で人の目を通して行ったものを、より効率的なアプローチ（話題の類似性の観点から）でより自動的に検出。
- 8月の情報処理学会（FIT2021）での発表で「奨励賞」を受賞。

3. この手法の課題

- 新聞記事になっているものしか見える化出来ないため、SDGsの目標によっては記事になりにくく分析が難しいものもある。

日本の力を、世界のために。

Supporting Your Global Challenges



3. まとめと今後の予定

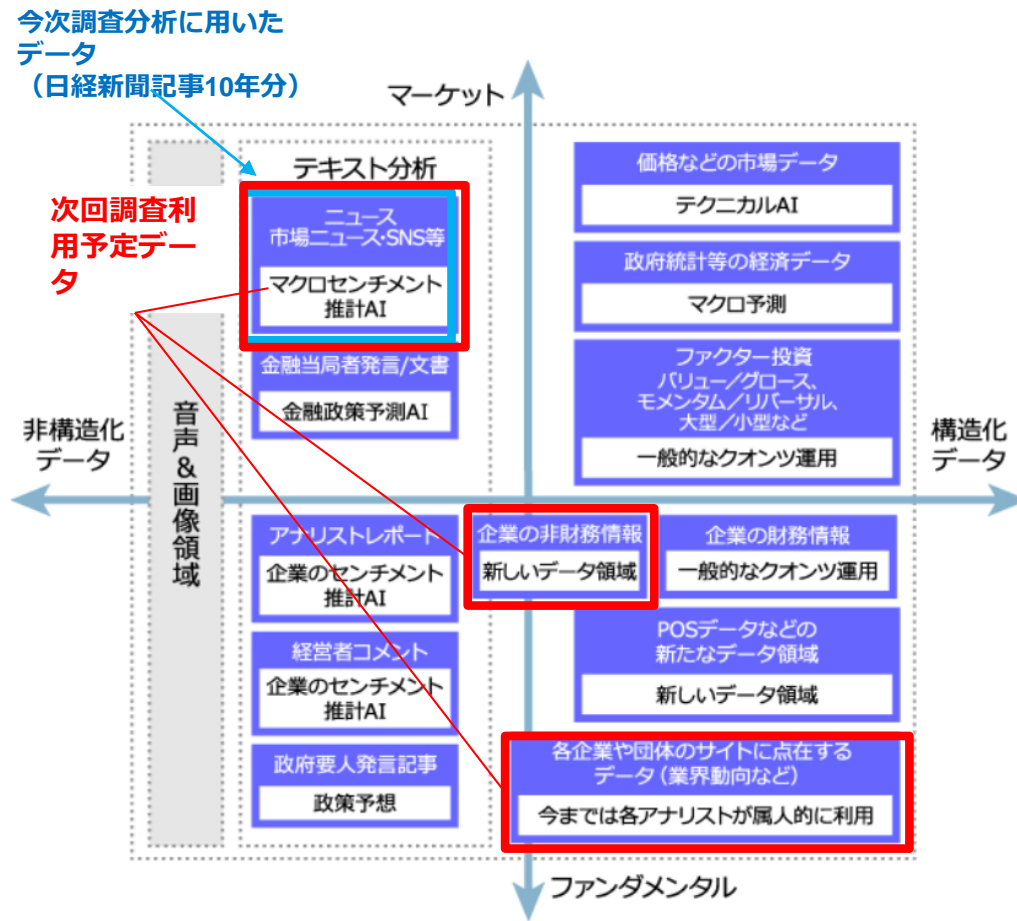


3. まとめと今後の予定

今回はいくつかの手法を試みた結果、「素早く気づきを与える分析手法」として、テキストマイニングの有用性が確認できた。他方、日本語のみを対象とした分析であったこと、新聞記事のみをデータソースとしたことなど、今後の拡張性については、更なる検討が必要。この辺をふまえ、今後の予定は以下。

◆ 今後の予定（10月～来年2月）

- ✓ **調査環境の改善**：ビックデータ分析・処理に対応可能な外部クラウドサービスの利用を予定。
- ✓ **データソースの大幅な拡充**：日経新聞の記事データに加え、今年度はAPIを経由したWebスクレイピング等の検索技術を活用し、英文記事データや国内外企業の非財務情報などデータソースを極力広げていく。
- ✓ **最新トレンドをテーマに選択**：今年度は脱炭素を巡る業界の事業戦略や技術動向など最新のトレンドを把握する。
- ✓ **対外発信能力の向上**：一部の調査項目についてはインハウスでの作業を実施予定。



株式会社国際協力銀行

企画部門 調査部 第1ユニット

春日 剛 (t-kasuga@jbic.go.jp)

成宮 仁 (j-narumiya@jbic.go.jp)

※意見交換して頂ける方を幅広く募集しております。

本資料は調査研究の参考資料として作成されたものであり、株式会社国際協力銀行の見解を表すものではありません。また、特定の商品・サービス等を推奨するものではありません。本資料の無断転用、公表等は固くお断りします。本資料の利用に際して損害が発生しても、弊行は一切の責任を負いかねます。