



# **BOJ** *Reports & Research Papers*

2015年6月

## ビッグデータを用いた経済・物価分析について

—— 研究事例のサーベイと景気ウォッチャー調査のテキスト分析の試み ——

日本銀行調査統計局

岡崎 陽介

敦賀 智裕

本稿の内容について、商用目的で転載・複製を行う場合は、予め日本銀行調査統計局までご相談ください。

転載・複製を行う場合は、出所を明記してください。

2015年6月  
日本銀行調査統計局

岡崎 陽介<sup>†</sup>

敦賀 智裕<sup>‡</sup>

## ビッグデータを用いた経済・物価分析について\*

—— 研究事例のサーベイと景気ウォッチャー調査のテキスト分析の試み ——

### ■要 旨■

近年、経済活動と情報技術（IT）の結び付きが強くなるもとで、大規模かつ多種多様なデータ（「ビッグデータ」）が生成され、経済・物価分析に徐々に活用されるようになりつつある。本稿では、同分野における2000年代後半以降の研究動向のうち、経済・物価動向の把握を目的としたものに注目する。そして、豊富な研究例が報告されている（1）インターネット検索データ、（2）POSデータ等のスキャンデータ、（3）テキストデータを用いたものを中心に、概要をとりまとめている。これらは、速報性・リアルタイム性の高さ、これまで定量化されてこなかった定性的な情報の活用、既存のマクロ経済指標を補完するような新たな経済指標の作成・公表等といった観点で、従来にはない視点を提供している。勿論、これら以外にも、カード決済や企業間取引等の多種多様なデータを用いた研究や、経済理論の構造的な実証や因果推論を目的とした研究など、適用範囲は急速に広がっている。

さらに、これまで研究事例は豊富に報告されてきたが、実務に応用されるケースが少なかったテキスト分析の応用可能性を評価するため、内閣府の「景気ウォッチャー調査」のテキストデータの分析を行った。その結果、テキストデータの機械的な処理に伴う日本語特有の難しさといった課題もみられたが、経済主体のセンチメントの特徴を評価する上で有用であることがわかった。今後は、こうした試みが研究機関や民間調査機関等においても行われていくことで、経済・物価分析に適した分析手法の確立、応用事例の蓄積等が一層進展していくことが期待される。

---

<sup>†</sup> E-mail: [yousuke.okazaki@boj.or.jp](mailto:yousuke.okazaki@boj.or.jp)

<sup>‡</sup> E-mail: [tomohiro.tsuruga@boj.or.jp](mailto:tomohiro.tsuruga@boj.or.jp)

\* 本稿の作成にあたっては、亀田制作、塩谷匡介、代田豊一郎、中島上智、中村康治、原尚子、前田栄治の各氏及び日本銀行のスタッフ各位から有益なコメントを頂いた。残された誤りは全て筆者に帰する。なお、本稿の内容と意見は筆者ら個人に属するものであり、日本銀行あるいは調査統計局の公式見解を示すものではない。

## 1. はじめに

近年、ヒトとモノの双方と情報技術（以下、IT）との結びつきが活発になり、世界に存在するデータの90%は過去10年に生成された、とも言われているように（Haldane[2013]）、膨大なデジタルデータの蓄積が急速に進んでいる。ヒトについては、1990年代後半以降、検索サイトやオンラインショッピングの利用が浸透したことに加え、2000年代に入ってからニュースサイトの記事配信サービスやソーシャルネットワークサービス（以下、SNS）、動画共有サイト等の普及が進んでいる。また、モノについても、従来からITとの結び付きが強いコンピュータ端末に加えて、家電、自動車、産業用機械等、様々なモノ同士がネットワークで接続され、データのやり取りを行う「IoT（Internet of Things）」が広がりを見せている。こうしたもとで、大規模かつ多種多様なデジタルデータが新たに次々と生成されており、これらを総称して「ビッグデータ」と呼ばれている<sup>1</sup>。

また、こうしたデータの活用を促進する技術革新も進行している。①従来は扱うことが難しかった大規模なデータや、テキストによる定性的な情報、音声・画像・映像データ、位置情報といった多様なデータの解析技術が発展・普及したこと、②そうしたデータの生成・収集を行う機器や、分析のためのソフトウェア、データベース、記憶媒体等の導入コストが低下したことも最近の大きな変化である。

こうしたもとで、ビッグデータを経済・物価分析に応用しようとする取り組みも、徐々に活発になってきている。ビッグデータの経済・物価分析への応用可能性としては、大きく2つの方向性から発展しており、①（既存統計を補い得るような）速報性の高さや、各種経済主体の経済活動を細かく捕捉することができるという特徴を活かして、経済・物価動向の把握を試みる方向性と、②経済理論の実証、とりわけ特定の政策が経済変数に与える効果や構造的な因果関係の実証に用いる方向性の研究が進められている（Einav and Levin[2014]）。

海外の主要中央銀行の取り組み内容をみると、ビッグデータ関連の研究論文の公表や、ワークショップの開催といった活動が行われ、①の経済・物価動向の把握を中心に調査研究が進められている様子が窺われる（図表1、2）。

---

<sup>1</sup> Bholat[2015]は、ビッグデータとは、①大規模、②高頻度、③多様式、などの要素のうち少なくとも1つ以上を備えているデータであるとしている。

【図表 1】 海外の主要中央銀行等での取り組み（1）

国・地域	取り組み	具体的な内容
米国	研究論文等の 公表	<b>FRB</b> ：2014年3月に公式ブログサイト「FEDS Notes」に、ニュース記事の <u>テキストデータ</u> を用いてセンチメント指標を構築し、金融市場との関係を分析する記事を掲載。また、これらの分析結果を研究論文として公表（Heston and Sinha [2014]等）。
		<p>ニューヨーク連銀：FOMC議事録の<u>テキストデータ</u>を用いて、各時点で特徴的なテーマ・概念を抽出し、足もとや将来の経済状況との関係を分析（Boukus and Rosenberg[2006]）。</p> <p>2012年1月には公式ブログサイト「Liberty Street Economics」に、<u>検索データ</u>を用いて、米国の住宅ローン借り換え申込者数等を推計する記事を掲載。</p>
カナダ	研究論文等の 公表	<p>カナダ中銀：中央銀行が公表する各種資料の<u>テキストデータ</u>を用いて、各時点で特徴的なテーマ・概念を抽出し、市場金利との関係について分析（Hendry and Madeley[2010]）。</p> <p>また、2013年夏季報にて、ビッグデータの特徴や意義、注意点等を紹介する記事を掲載。</p>
	ワークショップ 等の開催	<b>ECB</b> ：2014年4月にワークショップを開催。 <u>検索データ</u> を用いた米国CPIの推計事例（Koop and Onorante[2013]）等を発表。
ユーロ エリア	研究論文等の 公表	<b>イタリア中銀</b> ： <u>検索データ</u> を用いて、イタリアや米国の失業率について分析（D'Amuri[2009]、D'Amuri and Marcucci[2010]）。
		<b>スペイン中銀</b> ： <u>検索データ</u> を用いて、イギリスからのスペイン観光客数について分析（Artola and Galán[2012]）。
		<b>オランダ中銀</b> ：インターネットオークションサイト上に記載されている、欧州8か国の価格の <u>スキャンデータ</u> を用いて、一物一価の法則について分析（Maier[2005]）。

（資料）各中央銀行

【図表 2】 海外の主要中央銀行等での取り組み（2）

国・地域	取り組み	具体的な内容
英国	ワークショップ等の開催	英国中銀: 2014年7月にワークショップを開催。ビッグデータがもたらす社会経済の構造変化への中央銀行の対応について議論(Bholat[2015])。 2015年2月のコンファレンスでも、ビッグデータを含む新たなデータの活用と中央銀行の調査業務の関わり方について議論(Bank of England[2015])。
	研究論文等の公表	英国中銀: <u>検索データ</u> を用いて、英国の住宅価格指数・失業者数を分析 (McLaren and Shanbhogue[2011]) 。
	専門チームの発足	英国中銀: 足もとの経済状況をリアルタイムに把握するための先端分析ユニット (Advanced Analytics Unit) を2014年度に新たに発足。
その他	ワークショップ等の開催	<u>IFC (アービングフィッシャー委員会)</u> : 2014年の年次大会 (9月) やサテライトワークショップ (11月) で、信用情報データベース等のマイクロデータが公的統計を補完する可能性等について議論。 また、各中央銀行のビッグデータの利用状況について情報共有を図るウェビナーを10月に開催 (Bank for International Settlements[2015]) 。

(資料) 各中央銀行、BIS

そこで本稿では、経済・物価動向の把握を目的とした国内外の研究事例を中心に、特に研究事例が多く報告されている、①検索データ、②POS データ等のスキャンデータ、③テキストデータ、の3つの潮流に焦点をあて、主要な分析結果を紹介する。そのうえで、他のデータによる経済・物価動向の把握の事例や、経済の構造や因果関係の実証といったその他の目的に基づく分析等にも若干言及する。

さらに、実際の分析例として、内閣府の「景気ウォッチャー調査」の景気判断理由を記載したテキストデータを用いた分析を行い、これまで実務に応用されるケースが少なかったテキスト分析の応用可能性を展望する。

## 2. ビッグデータを用いた経済・物価分析の最近の動向

ビッグデータを用いた研究事例は急速に増加しているが、本稿で注目する経済・物価動向の把握を目的としたもののうち、特に研究が盛んなものをまとめると、次のような3つの潮流に大別することができる<sup>2</sup>。

- ① 検索サイトが提供する検索頻度データを用いて家計活動の分析を行うもの
- ② POS データ等のスキャンデータを用いて消費・物価動向の分析を行うもの
- ③ SNS やオンラインで配信される公的文書、企業の IR 資料や各種記事に含まれるテキストデータを用いて金融市場や企業・家計活動の分析を行うもの

以下では、経済・物価動向の把握を目的とした事例について、上記の3つのデータ分類ごとに、主要な研究報告の特徴・結果等を整理して紹介したうえで、それ以外のデータを用いた研究にも言及する。また、経済理論の実証・因果推論など、上記以外を目的とした分析事例も紹介する。

### 2-1 検索データの利用

2006年頃より大手検索サイト Google によって検索頻度データの提供が開始されて以降、自動車販売や住宅販売、旅行者数、失業率といった家計の経済活動の分析における、検索データの有用性を報告する事例が多くみられるようになっている (Edelman[2012]、白木他[2013])。具体的には、経済指標をナウキャスト (=同時点までに入手可能な情報を用いて、足もとの指標を公表前に推計) する際に、検索データを説明変数として推計式に投入することで説明力や予測力が向上したことを示す報告が中心となっている。

たとえば、McLaren and Shanbhogue[2011]では、それぞれ「estate agents」、「jobseeker's allowance (失業手当給付金)」をキーワードとした検索データを用いて、英国の住宅価格指数および失業者数についてのナウキャストを行い、説明力が向上したという結論を報告している (図表3)。

---

<sup>2</sup> これら3つの潮流の他にも多種多様なデータを用いた分析が提案されている(2-5項を参照)。

【図表3】 検索データを利用した経済・物価動向の分析事例

分野	推計対象
個人消費	米国の自動車販売や旅行者数 (Choi and Varian[2011])
	スペインの観光客数 (Artola and Galan[2012])
	日本の国内旅行取扱額 (白木他[2013])
住宅・不動産	米国の住宅販売 (Choi and Varian[2011])
	英国の住宅価格指数 (McLaren and Shanbhogue[2011])
	米国の住宅ローン滞納状況 (Askitas and Zimmermann[2011])
雇用	イタリアの失業率 (D'Amuri[2009])
	ドイツの失業率 (Askitas and Zimmermann[2009])
	米国の失業率 (D'Amuri and Marcucci[2010])
	英国の失業者数 (McLaren and Shanbhogue[2011])
その他	米国の消費者物価 (Guzmán[2011]、Koop and Onorante[2013])
	米国のマネーサプライ (Koop and Onorante[2013])

検索データが重要な意味をもつようになっている背景としては、消費や失業手当の受給といった家計活動が行われる際に、家計が検索サイトを通じて購入する財品目や価格、購入場所、公的手当て申請手続きの方法等の調査を伴う傾向が強まっていることが考えられる。また、これらの活動そのものがオンラインショッピングや電子手続きの普及により、インターネット内で行われる傾向が強まっていることも指摘できる。

こうした点を踏まえると、検索データを用いた分析の主な特徴は、①「検索行動」という行為を定量化し、家計の経済活動を数値的に分析すること、②リアルタイム性の高さを活かして、経済指標のナウキャスト・予測を行うこと、の2点に整理することができる。

なお、検索データを利用する際には留意点も存在する。たとえば、経済変動と必ずしも関わりのないような大きなニュースが発生した際には、検索頻度が意図しない変化を示すおそれがあり、ナウキャストや予測の精度悪化につながり得る。そのため、実際の実務に活用する際には、機械的に運用するのではなく、検索キーワードの選定等を含めて、客観情勢をふまえた柔軟な運用が必要だと考えられる。

## 2-2 スキャンデータの利用

第二の潮流として、POS データに代表されるスキャンデータを用いた分析が活発である。スキャンデータとは、実店舗やインターネット上の仮想店舗での販売時点・価格・販売数量等を機械的に収集することにより記録したデータを指す<sup>3</sup>。

特に目立った利用例として、スーパーやコンビニ等の POS データや、インターネット小売業者が Web 上に記載する価格等を機械的に収集（「スクレイピング」）することで得たデータ等を用いて、高頻度な価格指数や消費水準指数等の作成を試みた研究が行われている（図表 4）<sup>4</sup>。

また、こうした新たな指数を用いた分析事例として、高頻度な価格指数と消費者物価指数の持つバイアスなどを検証し、家計が直面する真の物価を示す物価指数を導こうとする研究（Handbury et al.[2013]等）や、家計の属性により直面する価格水準に不均一性が生じる背景を調査したもの（Abe and Shiotani[2014a、2014b]等）など、政府が公表する物価指数を補完するかたちで物価動向の把握に役立てようとする取り組みがみられる（図表 5）。

こうした点を踏まえると、スキャンデータを用いた分析の主な特徴は、リアルタイム性の高さを活かして、より足もとの経済情勢を分析・把握すること、特に、高頻度という強みを活かした補完的な指標を作成することにあると整理することができる。もっとも、既存の統計との動きを比較する際には、調査対象品目のカバレッジや品質調整方法、集計する際の算式等といった相違点に十分注意する必要がある。たとえば「東大渡辺研究室・日経日次物価指数（以下、日経・東大日次物価指数）<sup>5</sup>」や「SRI 一橋大学消費者購買指数」についてみると、図表 4 にある通り、現在のところカバレッジが CPI や家計調査の 2 割程度であるほか、CPI と比較を行う上では、集計する際の算式（両指数はトルンクヴィスト式である一方、CPI はラスパイレズ式）や、品質調整をどの程度実施しているかという点でも違いが存在<sup>6</sup>。そのため、単純な比較は困難となっている。

<sup>3</sup> POS のような実店舗から収集したデータを狭義のスキャンデータとし、インターネットから収集したデータと区別する研究も存在している。ここでは、両者とも機械的に収集・記録されている点に着目して、広義のスキャンデータと位置付けている。

<sup>4</sup> 図表 4 に挙げたものの他にも、スーパーマーケットおよびドラッグストアで使用されるポイントカードの購買履歴データをもとに、性別・年代別の属性ごとの物価指数の作成を試みるといった取り組みもみられる。

<sup>5</sup> 「東大渡辺研究室・日経日次物価指数」は、国立大学法人東京大学および日本経済新聞社の登録商標である。

<sup>6</sup> CPI と「日経・東大日次物価指数」の乖離要因については、丸山他[2015]を参照。

【図表4】 スキャンデータを用いた各種指数作成の取り組み

	東大渡辺研究室・ 日経日次物価指数 〔2013年5月公表開始〕	SRI一橋大学消費者 購買指数 〔2014年9月公表開始〕	Billion Prices Project Daily Online Price Index 〔2008年4月公表開始〕
公表 データ	<ul style="list-style-type: none"> <li>日次および月次系列を公表</li> <li>日本が対象</li> <li>1989/4/1日～</li> </ul>	<ul style="list-style-type: none"> <li>週次系列のみ公表</li> <li>日本が対象</li> <li>2014/9/30日週～</li> </ul>	<ul style="list-style-type: none"> <li>日次系列のみ公表</li> <li>世界22か国が対象（ホームページ上では米国とアルゼンチンを公表）。</li> <li>米国は2008/7/31日～、アルゼンチンは2007/12/1日～</li> </ul>
使用 データ	<ul style="list-style-type: none"> <li>POSデータ</li> </ul>	<ul style="list-style-type: none"> <li>POSデータ</li> </ul>	<ul style="list-style-type: none"> <li>インターネット小売業者がWeb上に記載する価格</li> </ul>
対象 店舗	<ul style="list-style-type: none"> <li>対象店舗は約300</li> <li>業態はスーパーマーケットのみ</li> </ul>	<ul style="list-style-type: none"> <li>対象店舗は約4,000</li> <li>業態はスーパーマーケットのほかGMSやコンビニエンスストア等</li> </ul>	<ul style="list-style-type: none"> <li>対象業者は数社</li> <li>業態はインターネット販売を提供する小売業者</li> </ul>
対象 商品	食料・飲料品 日用品雑貨等	食料・飲料品 日用品雑貨等	食料・飲料品 日用品雑貨、家電、 衣料品、家具等
	消費者物価指数の約17%	家計調査の約20%	

(注) 「東大渡辺研究室・日経日次物価指数」は、国立大学法人東京大学および日本経済新聞社の登録商標。

(資料) The Billion Prices Project、東大日次物価指数プロジェクト、流通・消費・経済指標開発プロジェクトの各ホームページより作成

## 【図表5】 スキャンデータを利用した経済・物価動向の分析事例

---

### (1) 公式物価指数とスキャンデータの比較

消費者物価指数（CPI）とPOSデータを用いて作成した指数を比較し、CPIのバイアスの存在を指摘（渡辺・渡辺[2013]、Handbury et al.[2013]）。

このほか、米国の代表的なオンライン書店がWeb上で提供する価格データを集計（スクレイピング）し、CPIに含まれる書籍の物価変動と比較研究したものの（Chevalier and Goolsbee[2003]）や、複数国のWeb上の価格データから作成した指数と各国のCPIを比較する研究も存在（Cavallo[2013]）。

---

### (2) 家計の属性を考慮した価格水準の分析

各家計が直面する価格水準には、所得や年齢階層、子供の人数等により不均一性があることをホームスキャンデータを用いて示した（Abe and Shiotani[2014a、2014b]）。

---

### (3) 新商品の情報を加味した単価指数の分析

小売価格の変動要因をPOSデータを用いて、①純粋な価格変化、②代替効果、③新商品導入効果の3つに分解した結果、消費増税後の単価指数は、③を主因としてCPIよりも高く推移していたことを示した（Abe et al.[2015]）。

---

なお、スキャンデータは高頻度に観察可能であるため、特売の影響<sup>7</sup>などについても分析可能というメリットを有する一方、基調を判断する上では注意を要する。たとえば、渡辺・渡辺[2013]によれば、1988年3月～2013年2月における「日経・東大日次物価指数」の前年比に関する標準偏差は1.09%（独立・同一分布を仮定して年率に換算すると20.8%）となっており、非常に大きい。マクロの基調的な変動を観察する際には、同指数に含まれる特売や暦の影響といったノイズ要因を識別し、基調成分を抽出する必要があると考えられる<sup>8</sup>。

## 2—3 テキストデータの利用

第三の潮流として、テキストデータに着目した分析が活発に行われている。この背景としては、近年、公的文書、企業のIR情報やニュース記事、SNSへの

---

<sup>7</sup> CPIでは、短期間の特売価格は、原則として調査しないこととなっている。ただし、特売が7日間以上続く場合には、特売価格を収集することもある。

<sup>8</sup> データのサンプリング頻度を下げて、月次のデータで標準偏差を計測すると、年率換算でも1.32%まで低下する。こうした、高頻度であるがゆえに特売や暦の影響によって生じるノイズを除去するため、渡辺・渡辺[2013]では7日間後方移動平均やHPフィルター等を用いた手法を試みている。

書き込み・口コミ情報といったデジタル化されたテキストデータが Web 上で容易に入手できるようになってきたことが指摘できる。また、分析手法（「テキストマイニング」）が発達したことや、分析ソフトウェアの導入コストが著しく低下したことも強く影響している。とりわけ自然言語処理の研究領域において、文章に含まれる語彙の品詞を機械的に識別して集計する「形態素解析（Morphological Analysis）」<sup>9</sup>の方法や、言葉がもつポジティブ／ネガティブな印象を表す「感情極性（Sentiment Polarity）」を数値化する研究成果が急速に蓄積されつつあり、経済・金融市場分析を含む幅広い領域に応用できるようになりつつある。

たとえば Li[2006、2010]や Loughran and McDonald[2011]、Davis et al.[2012]、Davis and Tama-Sweet[2012]等は、企業の IR 資料等に記載されているテキスト情報を用いてポジティブ／ネガティブな語彙の出現頻度を集計した企業マインド指標が、将来の企業収益に対して予測力を持つことを示している<sup>10</sup>（図表 6）。

【図表 6】テキストデータを利用した経済・物価動向の分析事例（1）  
— 各種マインド指標の抽出・作成 —

①消費者態度	<p>テキストデータに含まれるポジティブな語彙数とネガティブな語彙数を集計し、消費者マインド指標に対する説明力や予測力を報告。</p> <p>データとしては、SNSへの投稿記事（Daas and Puts[2013]）や、ブローカーレポートを使用（Nyman et al.[2014]）。</p>
②企業マインド	<p>企業の経営状態に関する経営者の定性的な見方や楽観的／悲観的な度合いを、Form10-KやMD&amp;A等の経営報告書のテキストデータを用いて指数化（Kogan et al.[2009]、Li[2010]、Loughran and McDonald[2011]、Davis et al.[2012]、Davis and Tama-Sweet[2012]、Brockman and Cicon[2013]）。</p>

<sup>9</sup> たとえば、「日銀太郎に会った。」という文章に対し、分析ソフトウェアが「日銀太郎（名詞）」、「に（助詞）」、「会っ（動詞、原形：会う）」、「た（助動詞）」、「。（句読点）」という形態素に機械的に判別する。また「景気動向」といった語については、「景気」＋「動向」という2つの語彙にわかち書きを行う。

<sup>10</sup> 企業マインドを抽出する研究はこの他にも、投資家向けカンファレンスコールの内容（Davis et al.[2014]）を対象としたものや、カンファレンスコールでの音声データを対象としたもの（Mayew and Venkatachalam[2012]）など、多岐にわたる。

また、テキスト情報から得られる経済主体のセンチメントや関心事の動向が、金融市場や経済指標と高い連動性を示すことを報告するものも多数みられる。

このうち、金融市場については、和泉[2012]や石島他[2013]、Heston and Sinha[2014]等が、ニュース記事や各種レポートで用いられる語彙について、感情極性対応表<sup>11</sup>を用いた重みづけを行って機械的に集計したセンチメント指標を構築する試みを行っている。何れも、同指標の変動が金融市場の騰落の方向と合致しているとの結果を報告している。また、SNS への投稿記事や Web 上のニュース記事等のテキストデータからより高頻度でリアルタイム性の高い分析を行ったものも存在する（迫村・和泉[2013]、迫村他[2013]、前川他[2013]等）（図表 7）。

## 【図表 7】 テキストデータを利用した経済・物価動向の分析事例（2）

### — 金融経済情勢とテキスト情報の連動性 —

#### ① 金融市場との連動性

新聞・雑誌等のニュース記事のテキストデータに含まれる語彙のトーンを集計したセンチメント指標を作成。同指標と株価との間に一定の連動性がみられた（Tetlock et al.[2008]、和泉[2012]、石島他[2013]、Heston and Sinha[2014]、沖本・平澤[2014]）。

SNS への投稿記事や Web 配信されたニュース記事等の、よりリアルタイム性の高いテキストデータによる同様の分析もある（迫村・和泉[2013]、迫村他[2013]、前川他[2013]）。

このほか、日本銀行が公表する金融経済月報から経済用語の出現パターンを集計し、各時点での文書の特徴的なテーマ・概念を抽出する試みや、市場金利の動向に対し説明力があることを示すものもみられた（三菱東京UFJ銀行[2007]、和泉他[2008、2011]）。

#### ② 経済指標との連動性

SNS 上の投稿データから失業に関する複数のキーワード（「lost my job」等）の出現頻度を計測し、それらの主成分から「ミシガン大学ソーシャルメディア失業インデックス」を作成。政府の新規失業保険申請件数の早期推定に有用であることが示された（Antenucci et al.[2014]）。

このほか、Fed の地区連銀経済報告における全米景気判断表現のトーンを分析し、ポジティブな場合は GDP の短期的変動に関連していた一方、ネガティブな場合は循環的変動に関連していたとの報告もある（Armesto et al.[2007]）。

<sup>11</sup> 「感情極性対応表」とは、各語彙がどの程度ポジティブ／ネガティブであるかを数値化し、対応付けを行ったもの。

経済指標については、米ミシガン大学の研究者らが、SNS への膨大な投稿データから、失業に関連する記事の出現頻度を計測し、新規失業保険申請件数の早期推定等に役立つ指標「ミシガン大学ソーシャルメディア失業インデックス」を開発している (Antenucci et al.[2014])<sup>12</sup>。

テキストデータを用いた分析手法をビジネスに応用しようとする取り組みも行われている。たとえば、海外の新聞社では、自社が提供する膨大なニュース記事データベースのテキストデータを用いて、主要国における月々のニュース・センチメント指標を算出し、顧客へ提供するサービスを行っている (図表 8)。

**【図表 8】 テキストデータを利用した民間企業の取り組み**

企業	取り組み内容
海外の新聞社 A社	自社が提供する膨大なニュース記事データベースのテキストデータを用いて、主要国における月々のニュース・センチメント指標を算出し、顧客へ提供。
海外の金融・経済分析 ソフトウェア開発会社 B社	FRBやECB、英国中銀、日本銀行等の各中央銀行の公表資料のテキストデータを用いて現状および先行きの景気に対する各中銀の強気/弱気度合いを算出するソフトウェアを開発。ヘッジファンド等に販売。
海外の投資家向け 情報提供会社 C社	米国企業のニュース記事のテキストデータを用いて市場センチメントを抽出。当該企業の株価動向に与える影響について投資家等へ情報を提供。
国内のデータ分析・ コンサルティング会社 D社	SNS上の株式関連の投稿データを用いて、株式市場に対する人々の強気/弱気度合いを定量化したセンチメント指標を投資家等へ提供。
国内の信託銀行 E社	ニュース記事のテキストデータから抽出した投資家心理についての情報を活用したファンドを運用予定。
国内の研究機関 F社	ニュース記事のテキストデータを用いて、企業の株価や信用力に対するポジティブ/ネガティブ度合いを抽出するソフトウェアを開発。資産運用や与信管理を行う顧客へ販売。

(資料) 各社プレスリリース等より作成

<sup>12</sup> 英語名称は、The University of Michigan Social Media Job Loss Index。ホームページ上で公表している (<http://econprediction.eecs.umich.edu/>)。

このような点を踏まえると、テキストデータを用いた分析の特徴は、①「文章」という定性的な記述内容から、経済主体のコンフィデンスや金融市場のセンチメントを数量化し、それらの動向に関する分析を行うこと、②データの頻度やリアルタイム性の高さを活かして、足もとの金融経済情勢の動向を分析・把握すること、の2点に整理できる。

なお、テキストデータを用いた分析を行う際の注意点としては、頑健性の問題が挙げられる。具体的には、対象とするテキストデータや抽出語彙の選択、わかち書きの処理精度等によって異なった結果となり得る<sup>13</sup>。したがって、分析目的に応じてヒアリング等のその他の手段を併用することが必要である。

また、センチメント指標の構築にあたっては、各語彙の評価の仕方によって結果が左右され得るという面があり、評価語彙のカバレッジ拡充や文脈に応じた評価の付与など<sup>14</sup>、今後の更なる進展が期待されている。

#### 2—4 ビッグデータを用いた経済・物価動向の分析のまとめ

ここまで、ビッグデータを用いて経済・物価動向を把握することを目的とした分析について、①検索データ、②スキャンデータ、③テキストデータの3つの潮流ごとに、主な分析結果を紹介してきた。これらは、以下のように整理できる（図表9）。

まず、3つのデータに共通している特徴として、従来より速報性・リアルタイム性の高い分析が可能であるという点を挙げることができる。これらの中には、足もとの経済・金融情勢の迅速な把握等を目指すものが多く含まれている。

次に、①検索データおよび③テキストデータを用いた分析では、「検索行動」や「文章」という定性的な情報を定量化して活用するという特徴が共通している。これは、従来は得られなかった、ないしは定量的に扱うことが困難であっ

---

<sup>13</sup> たとえば、「消費税増税」という記述を「消費税+増税」とするかどうかで結果は異なり得る。「円安ドル高」についても同様に「円安+ドル高」とするかどうかで違いが生じる。

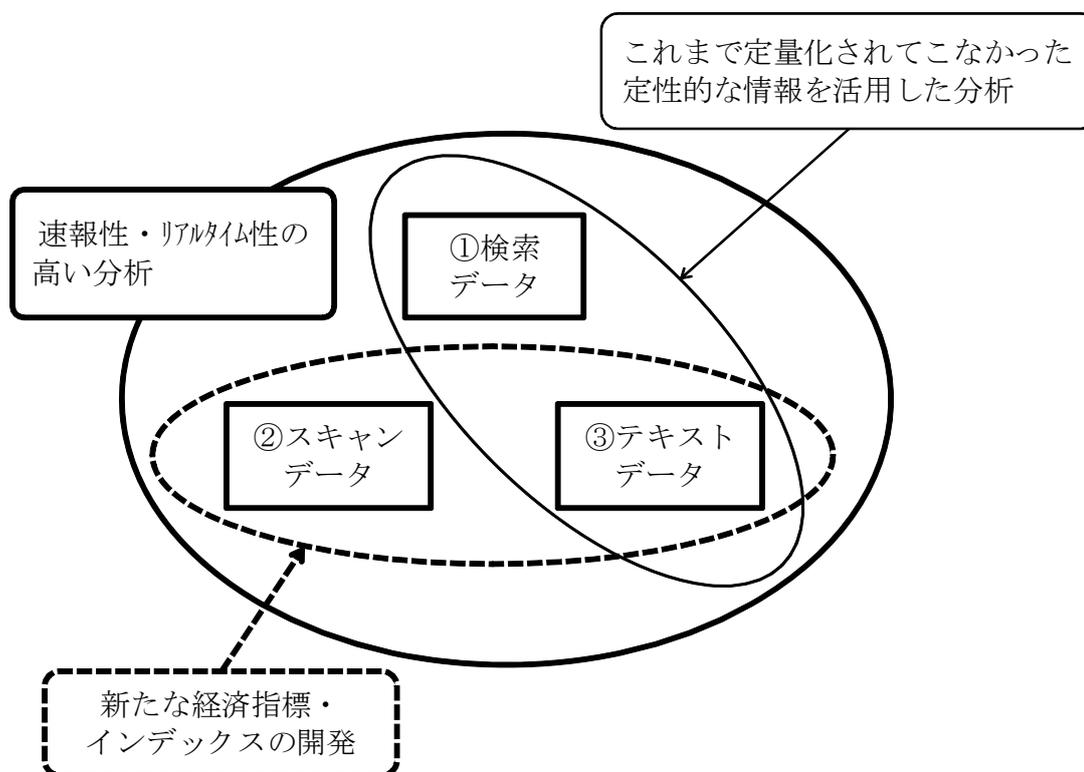
<sup>14</sup> Heston and Sinha[2014]は株価収益率の検証を行う際、感情極性対応表の選択により、予測力は異なったとの結果を報告している。

たとえば「消費」という語彙は、物が無くなるという印象を含んでおり（例：「エネルギーを消費する。」等）、一般的な文脈ではネガティブな感情を想起させ得るが、経済分析においては、需要の増加を示すポジティブな印象に着目したいことが多い。英語では既に経済・市場動向分析の文脈に即した感情極性対応表が作成され、無償公開されているが、対応している語彙数が少ないという別の問題がある。日本語においても経済分析の目的に即した感情極性対応表が作成・拡充されていくことが望まれる。

た情報を利用して分析を行おうというものである。

さらに、これらの一環として、既存のマクロ経済指標を補完するような新たな経済指標・マインド指数の作成・公表といった活動が行われている<sup>15</sup>。既述のように、速報性やリアルタイム性、定性的な情報を活用しつつ、経済・物価の動向を定点観測することができるといったメリットがある。

【図表 9】 経済・物価動向の把握のためのビッグデータ利用



## 2—5 ビッグデータを用いた分析の広がり

上記では、経済・物価動向の把握を目的とした分析を中心に、海外の中央銀行などで特に活発に研究が進められている3つの潮流を軸にまとめた。もっとも、第1節で述べたとおり、ビッグデータを用いた経済・物価分析への応用は、

<sup>15</sup> 経済産業省[2015]では、検索データやPOSデータ、位置情報データ等の公的統計や新たな経済指標づくりへの活用可能性を検討している。

これらのみに留まらない。以下ではその点について簡単に触れることとする。

まず、これら以外にも多種多様なデータを用いた研究がなされている。たとえば、クレジットカード等の決済件数データを用いて GDP をナウキャストしたもの (Galbraith and Tkacz[2013]) や、わが国の企業間取引データから商流構造を把握する取り組み (中小企業庁[2014])、地理情報システム (GIS) から得られる空間データを用いて東日本大震災の被害額の推定を行ったもの (斎藤他[2014])、企業の投資家向けカンファレンスコールの音声データから企業マインドの把握を試みたもの (Mayew and Venkatachalam[2012]) 等、枚挙にいとまがない程、多種多様なデータが活用されている<sup>16</sup>。

次に、経済・物価動向の把握以外の用途、とりわけ経済理論の実証分析や、特定の政策効果・因果関係の分析を行ううえでビッグデータを活用する動きもみられている。先のスキャンデータを用いた例でいえば、価格粘着性に関する実証 (Mizuno et al.[2010]、Cavallo and Rigobon[2013]等)、価格変動特性の国際比較 (Cavallo et al.[2013]) といった価格変動に関する研究や、オークションサイトで観察される消費者行動に与える税制の影響に関する研究 (Einav and Levin[2014]) 等が行われている。また、テキストデータについても、一部ではあるが、企業や産業ごとのビジネスの特徴抽出に焦点を当てた研究が行われており、たとえば Hoberg and Phillips[2010]は、買収・合併を行った企業の Form 10-K に含まれる製品説明のテキストデータを用いて、製品類似度を計測し、両企業が属する製品市場の類似性が高いほど、合併後の株価収益率やキャッシュフロー等が大きくなる「シナジー効果」が強くみられることを実証している。

さらには、複数の個票データを組み合わせることにより従来になかったようなミクロの分析手法を提供するものが現れてきている。たとえば、Chetty et al.[2014]は、約 4000 万件もの納税記録データを基に、所得格差が親子世代間でどのように変化するかを検証した。その際、同データに含まれる親と子の所得データをトラッキングして紐づけるのみならず、子の居住地の郵便番号を用いて国勢調査から得られる当該地域の人種構成を示すデータや、所得の格差を示すデータ、全米教育統計センターが提供する一人あたり教育費の水準に関する個票データを紐づけ、所得格差の変化に作用を及ぼし得る要因を検証している。また、Abowd et al.[1999]も、同様の方法で個人の所得データと人口動態データに含まれる教育年数、および雇用主企業の個票データを結びつけ、所得に影響を及ぼす要因等について報告している。

---

<sup>16</sup> 内閣府[2015]では、ビッグデータの現状とその活用可能性についてまとめており、その中で貨物や旅客の属性・移動経路等といった交通・運輸に関するデータの経済動向把握への活用可能性を検討している。

このように、ビッグデータを用いた経済・物価分析は、対象となるデータの種類と、その活用方法の両面で相当な広がりが見られ、潜在的な分析対象範囲は飛躍的に拡大していると考えられる。勿論、個人情報保護の観点から、情報の秘匿性確保等といった重要な課題も指摘されており、議論が進められている。

### 3. テキストデータを用いた景気ウォッチャー調査の分析事例

ここでは、これまで研究事例は豊富に報告されてきたが、実務に応用されるケースが少なかったテキスト分析の応用可能性を確認するため、代表的なマインド指標である内閣府の「景気ウォッチャー調査」に含まれる景気の現状および先行きに関する判断理由集のテキストデータを用いて、その記述内容の特徴変化を定量的に観察する。リーマンショック後の近年のマクロ経済理論では、不確実性ショックの影響を強調するものが多くみられ（たとえば Bloom[2009]）、日々の経済主体のマインドがどのように変化しているかを把握することは有益である。

#### 3-1. データの概要と分析アプローチ

「景気ウォッチャー調査」は、各地域の景気動向を迅速に把握するために、地域の景気に関連の深い動きを観察できる立場にある人々（景気ウォッチャー）の景気判断を調査する目的で、内閣府が毎月作成・公表している経済調査である。月々の景気判断を集計した DI の他に、全国の景気ウォッチャーが景気の現状と先行きを判断した事由を家計・企業・雇用動向のそれぞれについて箇条書きした「景気判断理由集」のデータが提供されている。これは、毎月計 120 ページ、センテンス数は 4 千以上、語彙数にして 10 万語以上に及ぶ豊富なテキスト情報を含んでいる。以下では、これを用いた分析を行う。

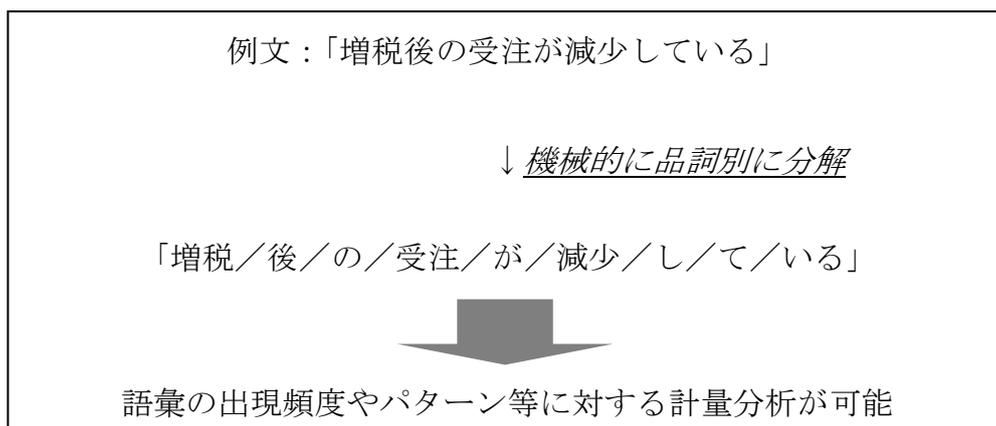
具体的な分析アプローチとしては、第一段階として、テキストデータを品詞別に分解（形態素解析）し、各語彙の出現回数を機械的にカウントする。こうすることで、文章中での各語彙の出現頻度やパターン等について、計量的な分析を行うことが可能となる（図表 10）。

これを用いて、第二段階として以下の 2 つの分析を行う。まず、①同理由集に含まれる語彙のポジティブ／ネガティブ度合いを集計したセンチメント指標を作成し、同調査で示される DI との対応関係を調べる。また、②文章中での出

現パターンの類似度（共起関係<sup>17</sup>）が強い語彙同士を線で結んだ共起ネットワーク図を作成し、同理由集の記述内容の特徴変化を分析・考察することで、経済主体のマインドに変化を与えた要因・トピックの特定を試みる<sup>18</sup>。

【図表 10】形態素解析の例と 2015 年 3 月調査（現状）での語彙の出現頻度

（形態素解析の例）



（2015 年 3 月調査＜現状＞での出現頻度上位 30 語）

<語彙>	<出現頻度> (%)	<語彙>	<出現頻度> (%)	<語彙>	<出現頻度> (%)
売上	1.63	多い	0.88	今	0.58
販売	1.39	減少	0.83	影響	0.57
需要	1.30	単価	0.80	傾向	0.56
良い	1.12	企業	0.76	変わる	0.56
増税	1.07	受注	0.70	価格	0.55
消費税	1.00	求人	0.69	小売	0.53
動き	0.99	景気	0.68	伸びる	0.53
増える	0.95	製造	0.68	悪い	0.52
駆け込み	0.90	続く	0.68	百貨店	0.51
増加	0.89	商品	0.60	今年	0.49

（注）語彙については「する」、「ない」、「なる」等の平仮名のみを除外して集計。

（資料）内閣府「景気ウォッチャー調査」より作成

<sup>17</sup> 「共起」とは、特定の文中において、ある語彙 A に着目した際に、別のある語彙 B が頻繁に出現することを表し、語彙 B の出現頻度が高いほど、語彙 A と語彙 B の共起関係は強くなる。

<sup>18</sup> ここでは、「KH Coder」（<http://khc.sourceforge.net/dl.htm> よりダウンロード）という分析ソフトウェアを用いている。

### 3-2. センチメント指標を用いた分析

まず、形態素解析を行って処理したテキストデータを用いて、センチメント指標を作成する。ここでは、センチメント指標の定義として、「感情極性対応表<sup>19</sup>」を用いて各語彙のポジティブ／ネガティブ度合いを表すウエイトを作成し、出現頻度で加重和したものとした（センチメント指標＝ $\sum_{\text{語彙}i} (\text{語彙}i\text{の感情極性}) \times (\text{語彙}i\text{の出現頻度})$ ）<sup>20</sup>。

景気の現状に関する判断理由集から作成した上記指標の動きをみると、ほぼ内閣府の「景気ウォッチャー調査」のDIと同様の動きとなっており、同理由集の記述内容が経済主体のマインド変化と密接に関連していることが確認できた（図表11）。

### 3-3. 共起ネットワーク図を用いた分析

次に、経済主体のマインドに変化を与えた背景を分析・考察する。まず、2014年12月から2015年3月にかけてセンチメント指標が大きく改善した要因を検証すると、感情極性が大きなプラス（マイナス）の値をとる「良い」（「悪い」）等の直截的な語彙の寄与率が大きく、これらの出現頻度に大きく左右されていることが確認できる（図表12）。そこで以下では、何が良いのか（悪いのか）を特定するために、共起ネットワーク図を作成し、各語彙がどのような文脈で用いられているかを確認していく。

共起ネットワーク図とは、語彙同士の共起関係の強弱を視覚化した図であり、これを用いて経済主体のマインドに変化を与えた要因・トピックを考察する。語彙同士の共起関係の強弱を測る尺度としては、Jaccard係数を用いる。Jaccard係数とは、語彙Aと語彙Bがあったとき、「AとBが同時に出現する頻度」と「AまたはBが出現する頻度」の比率として定義される指標であり、語彙Aと語彙Bの共起関係が強く、特定の文中において同時に出現することが多ければ、Jaccard係数は大きくなる性質を持つ（語彙Aと語彙BのJaccard係数＝「AとBが同時に出現する頻度」／「AまたはBが出現する頻度」）<sup>21</sup>。

<sup>19</sup> ここでは、石島他[2013]に倣い、高村他[2006]の研究に基づいた表を高村氏のホームページ（[http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html)）よりダウンロードして使用した。なお、同対応表は必ずしも経済分析の目的に即したものではないため、分析結果については幅を持ってみる必要がある。

<sup>20</sup> 石島他[2013]では、各語彙がポジティブであれば+1、ネガティブであれば-1のウエイトで加重和したものをセンチメント指標の定義としている。

<sup>21</sup> ここでは、Jaccard係数を計算するテキストの範囲として、1つ1つの文を集計単位とし

【図表 1 1】センチメント指標の作成方法と推移

(作成方法)

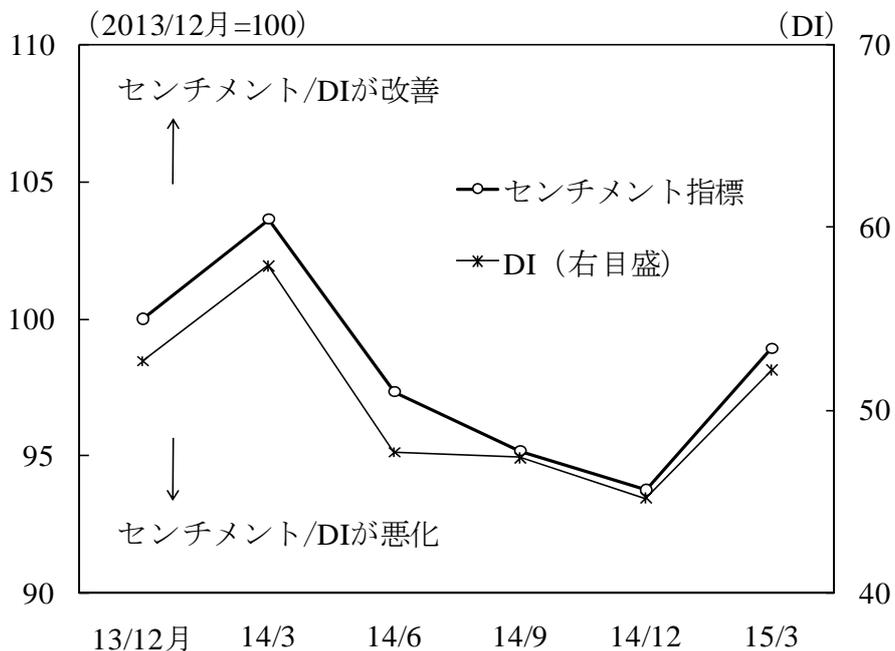
$$\text{センチメント指標} = \sum_{\text{語彙}i} (\text{語彙}i\text{の感情極性}) \times (\text{語彙}i\text{の出現頻度})$$

- ・「感情極性」とは、各語彙のポジティブ／ネガティブ度合いを数値化したもの  
(感情極性の例)

<語彙>	<度合い>
増税	-0.1
後	-0.4
受注	-0.1
減少	-0.4

⇒例文 (図表 1 0) のセンチメント指標：  
 $(-0.1) \times 1/4 + (-0.4) \times 1/4$   
 $+ (-0.1) \times 1/4 + (-0.4) \times 1/4 = \underline{-0.25}$

(推移)



(資料) 内閣府「景気ウォッチャー調査」

た。また、図表 1 3、1 5、1 7 では、出現頻度を上位 70 語程度に絞り込んだ上で、同係数が 0.06 以上の共起ネットワーク図を作成した。

【図表 1 2】感情極性が大きなプラス／マイナスをとる主な語彙の  
出現頻度とセンチメント指標の改善に対する寄与率

＜語彙＞	＜感情極性＞	＜出現頻度の変化＞ (14/12月→15/3月、%P)	＜センチメント指標改善に対する寄与率＞ (%)
良い	+1.0	+0.24	11.0
好調	+1.0	+0.08	3.8
意欲	+1.0	+0.07	3.3
期待	+1.0	+0.03	1.4
効果	+0.9	+0.07	2.9
減少	-0.4	-0.38	6.3
影響	-0.6	-0.62	18.0
ない/ぬ (否定助動詞)	-1.0	-0.29	13.3
厳しい	-1.0	-0.13	6.0
悪い	-1.0	-0.20	9.4
(合計)			75.4

(資料) 内閣府「景気ウォッチャー調査」より作成

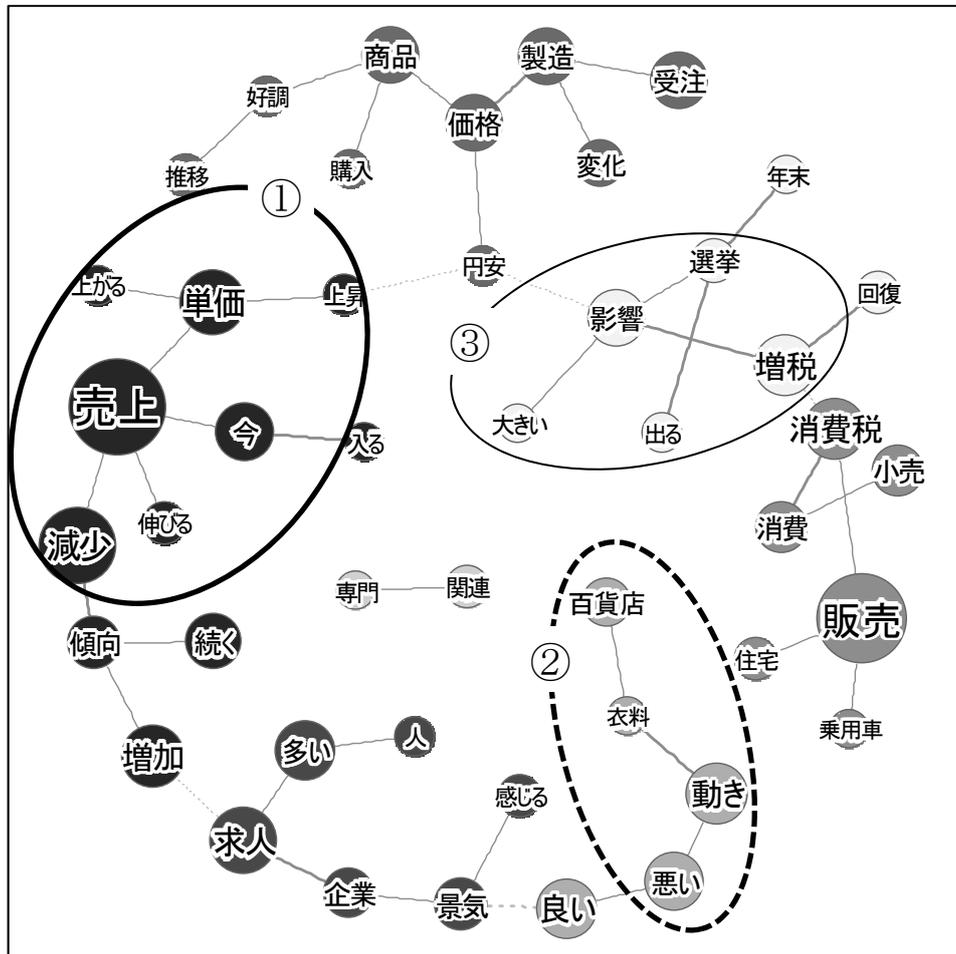
### (1) 2014年11月調査(現状)の共起ネットワーク図を用いた分析

まず、DIがこの期間でもっとも低かった2014年11月調査における景気の現状判断に関する判断理由集のテキストデータを用いて描いた共起ネットワーク図を観察する(図表13)。図のバブルの大きさは各語彙の登場頻度を表し、それぞれをつなぐ線の太さが共起関係の強さを表している<sup>22</sup>。

DIが悪化した要因を分析・考察するため、大きなマイナスの感情極性を持つ①「減少」や②「悪い」、③「影響」を含むネットワークに注目すると、①では「減少」－「売上」－「単価」－「上昇」・「上がる」といった共起関係が、②では「百貨店」－「衣料」－「動き」－「悪い」の共起関係が、③では「大きい」－「影響」－「増税」等の共起関係がそれぞれ観察可能。より仔細にみると、同理由集の中には、上記の語彙同士が同時に出現する記述が多くみられた(図表14)。こうした点を踏まえると、2014年11月調査でDIが悪化した背景として、単価が上昇する一方で数量要因により売上高が減少したことの影響や、天候要因の影響で百貨店での冬物衣料品販売の動向が低調であったこと、消費税引き上げの影響が引き続きみられているといったことが示唆される。

<sup>22</sup> バブルの色については、比較的強い共起関係にある一連の語彙グループを同色としており、色の濃淡に特段の意味は無い。

【図表 1 3】 2014 年 11 月調査（現状）の共起ネットワーク図



(資料) 内閣府「景気ウォッチャー調査」より作成

## 【図表 1 4】2014 年 11 月調査（現状）の判断理由集の主な記述

- 
- ①客単価が上昇する一方、売上全体は「減少」していることを指摘する記述
- ・ 販売単価は上昇傾向にあるものの販売点数の減少により、トータルの売上高は減少となっている（住関連専門店・経営者）
  - ・ 宿泊単価はバブル崩壊以降では最高となっているが、利用客数が減少したため、売上高は微減となっている（観光型旅館・スタッフ）
- 
- ②百貨店での衣料品販売が低調であったことを指摘する記述
- ・ コート類の動きが鈍いなど冬物の衣料品の動きは全体的に低調である（百貨店・経営者）
  - ・ 気温が若干平年より高いためかコート等の防寒衣料品の動きが悪い（百貨店・経理担当）
- 
- ③消費税引き上げの影響を指摘する記述
- ・ 消費税増税の影響から未だに回復してこない（スーパー・店長）
  - ・ 消費税増税の影響は薄まってきているものの、販売点数の落ち込みが止まらない状況（スーパー・総務担当）
- 

（資料）内閣府「景気ウォッチャー調査」より作成

## （2）2015 年 3 月調査（現状）の共起ネットワーク図を用いた分析

次に、経済主体のマインドが改善した局面として、2015 年 3 月調査における景気の現状判断に関する判断理由集のテキストデータを用いて描いた共起ネットワーク図を観察する（図表 1 5）。

これをみると、2014 年 11 月調査に引き続き、「上昇」－「単価」－「売上」－「減少」といった共起関係がみられているが、①「景気」－「良い」・「回復」といった共起関係が現れており、「良い」のバブルの大きさが大きくなっている。これは、同月の DI が 50 を上回り、景気が回復基調にあることと整合的である。また、大きなプラスの感情極性を持つ②「好調」という語彙を含むネットワークに注目すると、「観光」との共起関係が観察でき、良好な観光需要が窺われる（図表 1 6）。ただし、その他の特徴点として、③季節的な受注増加を指摘する見方（「受注」、「建設」、「年度末」等の共起関係）や、④前年に消費増税前の駆け込み需要がみられたため、それと比較して判断を保留ないし慎重化する見方（「消費税」、「昨年」、「駆け込み」等の共起関係）が観察されることから、同月の DI の評価にあたっては幅を持つ必要がある。



## 【図表 1 6】2015 年 3 月調査（現状）の判断理由集の主な記述

---

### ①景気が回復基調にあることを指摘する記述

- ・ 来客数が前々年並みに戻ってきおり、少しずつ景気が上向いていると実感している（衣料品専門店・店長）
- ・ ムード的には景気が良くなっている感覚があり、今までにない新製品の企画の話も少し増えている（窯業土石製品製造業・経営者）

---

### ②観光需要の好調さを指摘する記述

- ・ 中国、台湾からの観光目的の客の利用が好調である（観光型ホテル・スタッフ）
- ・ 3月に北陸新幹線開業があり、当地区の飲食店では観光客が多く来店している（商店街・代表者）

---

### ③季節的な受注増加を指摘する記述

- ・ 年度末の予算消化的な受注が増加している（窯業土石製品製造業・社員）
- ・ 年度末という季節要因はあるが、受注は増加している（建設業・経営者）

---

### ④昨年の増税前の駆け込みによる影響に言及する記述

- ・ 昨年の消費税増税前の駆け込み需要を念頭に、今年の実績をどうみるかも意見が分かれ、実勢の判断が難しい（百貨店・商品担当）
- ・ 消費税増税前の駆け込み需要があった昨年に比べ売上は18%減であるが、一昨年比では4.5%減とほぼ想定の範囲内（百貨店・売場マネージャー）

---

（資料）内閣府「景気ウォッチャー調査」より作成

## （3）2015 年 3 月調査（先行き）の共起ネットワーク図を用いた分析

最後に、2015/3月調査における景気の先行き判断に関する判断理由集のテキストデータを用いて描いた共起ネットワーク図を観察する（図表 1 7）。



## 【図表 18】2015年3月調査（先行き）の判断理由集の主な記述

### ①景気回復基調の継続を指摘する記述

- ・ 工場の増設や人員の増強を考えているところが多く、景気が良くなる傾向はしばらく続くとみている（電気機械器具製造業・営業担当）
- ・ 採用に積極的な企業が多くみられ、このまま景気の良い状態が続くのではないかとみている（人材派遣会社・社員）
- ・ 突発的な不安材料が出ない限り、現在の良好な景気が暫くは続くものとみている（百貨店・販売促進担当）

### ②先行きのベースアップに言及する記述

- ・ 今後、ベースアップ等での所得回復によって消費の回復が期待できる（百貨店・後方担当）
- ・ 4月以降の賃金のベースアップによる個人消費の底上げを期待している（都市型ホテル・スタッフ）
- ・ 大手のベースアップなどの情報はあがるが、地方企業はまだまだ給与面などで厳しい状態が続く（スーパー・企画担当）
- ・ 大企業でのベースアップがニュースになっているが、中小企業ではなかなか難しく、現状維持がやっとである（会計事務所・社会保険労務士）

（資料）内閣府「景気ウォッチャー調査」より作成

このように、経済主体の景気に対するセンチメントの動きと、その背景に関する定性的な情報を対応させて把握できる点は本手法の大きなメリットである。

## 4. 今後の展望と課題

本稿では、ビッグデータを用いた分析のうち、経済・物価動向の把握を目的とした分析事例について、①検索データ、②スキャンデータ、③テキストデータ、という3つの潮流を中心にまとめた。その結果、速報性・リアルタイム性の高さや、従来は定量化が困難であった定性的な情報へのアクセス可能性といった有用性を確認することができた。また、こうした分析の対象範囲が急速に拡大していることを確認した。

確かに、データ作成方法や集計方法に応じて、ノイズや季節性といった独特

の変動パターンを適切に処理する必要があることや、ビッグデータを用いて作成した指標と既存統計の比較においては、カバレッジや定義の違いを踏まえて解釈する必要があることなど、今後の課題や留意点はあるが、分析手法や集計方法等に関する一層の技術革新を積み重ねつつ、継続的な調査・研究を行っていくことで、ビッグデータが持つ可能性を引き出すことができると考えられる。

最後に、ビッグデータの整備・利活用が円滑に進められていくためには、データの二次利用や、機械可読性の高い形式での利用が可能な環境が必要である。また、データが持つ個人のプライバシー保護や、データを保有する企業の情報セキュリティ確保といった問題も同時に発生する。こうした点を十分に踏まえたうえで、データの有効活用が図られていくことが望まれる。

以 上

## 参考文献

- [1] 石島博・數見拓朗・前田章、「日次データを用いた市場センチメント・インデックスの構築と株価説明力の分析」、『第 11 回人工知能学会金融情報研究会資料』SIG-FIN-011-06、2013 年
- [2] 和泉潔、「大規模テキストを利用した経済指標分析手法に関する研究」、『学際大規模情報基盤共同利用・共同研究拠点 平成 24 年度共同研究 中間報告書』12-DA05、2012 年
- [3] 和泉潔・後藤卓・松井藤五郎、「テキスト情報を用いた金融市場分析の試み」、第 22 回人工知能学会全国大会 発表論文、2008 年
- [4] 和泉潔・後藤卓・松井藤五郎、「経済テキスト情報を用いた長期的な市場動向推定」、『情報処理学会論文誌』Vol.52 No.12、2011 年、3309-3315 頁
- [5] 沖本竜義・平澤英司、「ニュース指標による株式市場の予測可能性」、証券アナリストジャーナル 52(4)、2014 年、67-75 項
- [6] 経済産業省、「平成 26 年度 ビッグデータを利用した新たな経済指標・分析手法の動向に関する調査研究」、2015 年
- [7] 齊藤誠・中川雅之・顧濤・泉谷将登・岩佐丈・木澤諒平・武藤蔵・張瑩・中村京介、「GIS データに基づいた東日本大震災津波被害と原発事故の影響に関する推計手続きについて」、一橋大学大学院経済学研究科 Discussion Papers No. 2014-09、2014 年
- [8] 迫村光秋・和泉潔、「Twitter テキストマイニングによる経済動向分析」、第 9 回人工知能学会 ファイナンスにおける人工知能応用研究会 発表論文、2013 年
- [9] 迫村光秋・和泉潔・セーヨーサンティ、「Twitter テキストとネットワークの解析による経済動向分析」、第 10 回人工知能学会 ファイナンスにおける人工知能応用研究会 発表論文、2013 年
- [10] 白木紀行・松村浩平・松本梓、「景気判断における検索データの利用可能性」、『日本銀行調査論文』、2013 年
- [11] 高村大也・乾孝司・奥村学、「スピンモデルによる単語の感情極性抽出」、『情報処理学会論文誌』Vol.47 No.2、2006 年、627-637 頁
- [12] 中小企業庁、「2014 年版中小企業白書」、2014 年

- [13] 内閣府、「公的統計におけるビッグ・データの活用に関する調査研究」、2015年
- [14] 前川浩基・中原孝信・岡田克彦・羽室行信、「大規模ニュースデータと株価収益率の予測可能性について」、第10回人工知能学会 ファイナンスにおける人工知能応用研究会 発表論文、2013年
- [15] 丸山歩・嶋北俊一・落合牧子・上田聖、「CPI と東大物価指数の乖離の分析について」、総務省 統計研究所彙報 第72号 No.4、2015年、55-78頁
- [16] 三菱東京UFJ銀行、「テキストマイニング手法を用いた経済市場分析の試み～日銀金融経済月報を題材として」、『Focus on the Markets』 No.24 2007.12.28、2007年
- [17] 渡辺広太・渡辺努、「スキャナーデータを用いた日次物価指数の計測」、『CARF ワーキングペーパー』 CARF-J-094、2013年
- [18] Abe, Naohito, and Kyosuke Shiotani (2014a), "Who Faces Higher Prices? -An Empirical Analysis Based on Japanese Homescan Data-," *Asian Economic Policy Review*, 9 (1), 94-115.
- [19] Abe, Naohito, and Kyosuke Shiotani (2014b), "Analysis of Price Level Heterogeneity across Households Based on the Geary-Khamis Price Index," Hitotsubashi University Research Center for Price Dynamics Working Paper Series No.11.
- [20] Abe, Naohito, Toshiki Enda, Noriko Inakura, and Akiyuki Tonogi (2015), "Effects of New Goods and Product Turnover on Price Indexes." RCESR Discussion Paper Series, DP15-2.
- [21] Abowd, John M., Francis Kranarz, and David N. Margolis (1999), "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251-333.
- [22] Antenucci, Dolan, Michael Cafarella, Margaret Levenstein, Christopher Re, and Matthew D. Shapiro (2014), "Using Social Media to Measure Labor Market Flows," NBER Working Paper, 20010.
- [23] Armesto, Michelle T., Rubén Hernández-Murillo, Michael T. Owyang, and Jeremy M. Piger (2007), "Measuring the Information Content of the Beige Book: A Mixed Data Sampling Approach," FRB of St. Louis Working Paper,

2007-010B.

- [24] Artola, Concha, and Enrique Galán (2012), "Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches," Banco de Espana Occasional Papers, 1203.
- [25] Askitas, Nikolaos, and Klaus F. Zimmermann (2009), "Google Econometrics and Unemployment Forecasting," *Applied Economics Quarterly*, 55(2) , 107-120.
- [26] Askitas, Nikolaos, and Klaus F. Zimmermann (2011), "Detecting Mortgage Delinquencies," IZA Discussion Paper, 5895.
- [27] Bank of England (2015), One Bank Research Agenda, Discussion Paper, February 2015.
- [28] Bank for International Settlements (2015), Irving Fisher Committee on Central Bank Statistics, 2014 IFC Annual Report.
- [29] Bholat, David (2015), "Big Data and Central Banks," Bank of England Quarterly Bulletin No.2015 Q1, 55(1), 86-93.
- [30] Bloom, Nicholas (2009), "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623-685.
- [31] Boukus, Ellyn, and Joshua V. Rosenberg (2006), "The Information Content of FOMC Minutes," Available at SSRN 922312.
- [32] Brockman, Paul, and Jim Cicon (2013), "The Information Content of Management Earnings Forecasts: An Analysis of Hard versus Soft Information," *Journal of Financial Research*, 36(2), 147-174.
- [33] Cavallo, Alberto (2013), "Online and Official Price Indexes: Measuring Argentina's Inflation," *Journal of Monetary Economics*, 60(2), 152-165.
- [34] Cavallo, Alberto, Brent Neiman, and Roberto Rigobon (2013), "Currency Unions, Product Introductions and the Real Exchange Rate," NBER Working Paper, 18563.
- [35] Cavallo, Alberto, and Roberto Rigobon (2013), "The Distribution of the Size of Price Changes," NBER Working Paper, 16760.

- [36] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014), "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 129(4), 1553-1623.
- [37] Chevalier, Judith, and Austan Goolsbee (2003), "Measuring Prices and Price Competition Online: Amazon.com vs. BarnesandNoble.com," *Quantitative Marketing and Economics*, 1(2), 203-222.
- [38] Choi, Hyunyoung, and Hal Varian (2011), "Predicting the Present with Google Trends," Google Technical Report.
- [39] D'Amuri, Francesco (2009), "Predicting Unemployment in Short Samples with Internet Job Search Query Data," MPRA Working Paper, 18403.
- [40] D'Amuri, Francesco, and Juri Marcucci (2010), "'Google it!' Forecasting the US Unemployment rate with a Google Job Search Index," FEEM Working Paper No.31-2010.
- [41] Daas, Piet J.H., and Marco J.H. Puts (2014), "Social Media Sentiment and Consumer Confidence," ECB Statistics Paper Series, 5.
- [42] Davis, Angela K., Jeremy M. Piger, and Lisa M. Sedor (2012), "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language," *Contemporary Accounting Research*, 29(3), 845-868.
- [43] Davis, Angela K., and Isho Tama-Sweet (2012), "Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases Versus MD&A," *Contemporary Accounting Research*, 29(3), 804-837.
- [44] Davis, Angela K., Weili Ge, Dawn Matsumoto, and Jenny Li Zhang (2014), "The Effect of Manager-specific Optimism on the Tone of Earnings Conference Calls," *Review of Accounting Studies*, 1-35.
- [45] Edelman, Benjamin (2012), "Using Internet Data for Economic Research," *Journal of Economic Perspectives*, 26(2) 189-206.
- [46] Einav, Liran, and Jonathan Levin (2014), "Economics in the Age of Big Data," *Science*, 346.6210. 1243089.
- [47] Galbraith, John W., and Greg Tkacz (2013), "Nowcasting GDP: Electronic Payments, Data Vintages and the Timing of Data Releases," Scientific Series

(CIRANO), 2013s-25.

- [48] Guzmán, Giselle (2011), "Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations," *Journal of Economic and Social Measurement*, 36(3) 119-167.
- [49] Haldane, Andrew G. (2013), "Why Institutions Matter (More than Ever)," Bank of England, Speech at Centre for Research on Socio-Cultural Change (CRESC) Annual Conference, School of Oriental and African Studies, London, September 4, 2013.
- [50] Handbury, Jessie, Tsutomu Watanabe, and David E. Weinstein (2013), "How Much Do Official Price Indexes Tell Us about Inflation?," UTokyo Price Project Working Paper Series No. 019.
- [51] Hendry, Scott, and Alison Madeley (2010), "Text Mining and the Information Content of Bank of Canada Communications," Bank of Canada, Working Paper, 2010-31.
- [52] Heston, Steven L., and Nitish Ranjan Sinha (2014), "News versus Sentiment: Comparing Textual Processing Approaches for Predicting Stock Returns," Robert H. Smith School Research Paper. Available at SSRN 2311310.
- [53] Hoberg, Gerard, and Gordon Phillips (2010), "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis," *Review of Financial Studies*, 23(10), 3773-3811.
- [54] Kogan, Shimon, Dmitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith (2009), "Predicting Risk from Financial Reports with Regression," In Proc. NAACL Human Language Technologies Conference, Boulder, May-June 2009.
- [55] Koop, Gary, and Luca Onorante (2013), "Macroeconomic Nowcasting Using Google Probabilities," Paper presented at the ECB Workshop on using big data for forecasting and statistics, Frankfurt, 7-8 April 2014.
- [56] Li, Feng (2006), "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?," Available at SSRN 898181.
- [57] Li, Feng (2010), "The Information Content of Forward-Looking Statements in Corporate Filings—A Naive Bayesian Machine Learning Approach," *Journal*

*of Accounting Research*, 48(5), 1049-1102.

- [58] Loughran, Tim, and Bill McDonald (2011), "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, 66(1), 35-65.
- [59] Maier, Philipp (2005), "A 'Global Village' without borders? International price differentials at eBay," DNB Working Paper No. 44.
- [60] Mayew, William J., and Mohan Venkatachalam (2012), "The Power of Voice: Managerial Affective States and Future Firm Performance," *The Journal of Finance*, 67(1), 1-43.
- [61] McLaren, Nick, and Rachana Shanbhogue (2011), "Using internet search data as economic indicators," Bank of England Quarterly Bulletin No.2011 Q2, 51(2), 134-140.
- [62] Mizuno, Takayuki, Makoto Nirei, Tsutomu Watanabe (2010), "Closely Competing Firms and Price Adjustment: Some Findings from an Online Marketplace," *Scandinavian Journal of Economics*, 112(4), 673-696.
- [63] Nyman, Rickard, Paul Ormerod, Robert Smith, and David Tuckett (2014), "Big Data and Economic Forecasting: A Top-Down Approach Using Directed Algorithmic Text Analysis," Paper presented at the ECB Workshop on using big data for forecasting and statistics, Frankfurt, 7-8 April 2014.
- [64] Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy (2008), "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *The Journal of Finance*, 63(3), 1437-1467.