

【研究動向】

経済学史研究におけるテキストマイニングの導入*

——その手法と意義を中心に——

下 平 裕 之

I はじめに——テキストマイニングとは何か——

本稿は、近年人文・社会科学において急速に普及している「テキストマイニング」という分析手法について、その普及の背景と方法、経済学史研究へ適用する意義および実際の分析事例を示すことを通じて、経済学史への導入を進める際の1つの指針となることを意図している。まず本節においてテキストマイニングの定義および普及の背景・現状を述べ、次にII節では、経済学史研究者がテキストマイニングを導入する際に有用な各種の手法について概説する。III節においては社会科学におけるテキストマイニングの方法論に関する議論の説明を行い、最後にIV節において経済学史研究にテキストマイニングを導入する意義とその適用事例を示す。

テキストマイニングは広義には「テキストデータを計算機で定量的に解析して有用な情報を抽出するためのさまざまな方法の総称」(松村・三浦 2014, 1)と表現されるが、本稿においては経済学史研究における適用を視野に入れ、単なる定量的解析にとどまらない「テキストデータを、言語処理技術を用いて構造化データ・変数に変換し、それをもとに知識発見、仮説発見および仮説検証を行う手法」(喜田 2018, 8)という定義を採用する。

テキストマイニングの発展と普及には、以下のような背景がある。すなわち、従来(1990年代頃まで)の統計分析では、情報を得るための費用が高く、その情報を処理するための演算能力が低いという環境を前提としていた。その中で「できるだけ小さい情報量から、世界の姿を知ろうとする試み」(岡嶋 2006, 10)が統計分析であり、想定された法則の事後検証を得意としてきた。ところが2000年代以降になり、大量のデータ・情報が安価に入手可能で、その情報を処理する演算能力が圧倒的に高くなるという環境変化が生じた。また、プログラミング言語やアルゴリズムの改善による強力なソフトウェアの開発と流布も圧倒的な力を持った(Ignatow and Mihalcea 2017, 5)。対象となる情報の質が格段に向上し、その量や範囲が飛躍的に増大した結果、大量の

* 本稿は、日本学術振興会科学研究費補助金・基盤研究B「経済理論の大衆化から経済政策の形成へ：テキストマイニングを応用した実証研究」(課題番号15H03330, 2015-19年度)による研究成果の一部である。

情報を一括して短時間に処理できるようになり、分析の精度が格段に高まった。

テキストマイニングは、「自然言語処理」と「データマイニング」の混合様式であると考えられる。自然言語処理とは、日常用語を分類や検索によって計算機が処理しやすい表現に直すことによって、テキストデータから新しい知見を引き出そうとする手法である。具体的には、テキストという自然言語を文字・単語などの単位に分解し、形態素解析や構文解析等を用い数量化・構造化するデータとして変換する。

一方データマイニングは、「データ内の情報や意思決定に使用される知識を特定するために用いられるさまざまな手法」(喜田 2008, 26)と定義される。通常のデータマイニングの対象が数値や商品名といった表形式で集計された数値からなる「構造化データ」であるのに対し、テキストマイニングの対象は加工されていない生の文章である「非構造化データ」からなる。そのためテキストマイニングにおいては、データとしてのテキストを範疇ごとに分類したり可視化することが必要となるのである。具体的には、データマイニングの様々な手法(頻度分析、共起分析、多変量解析等)を適用するとともに、グラフや図表等によって可視化する作業を行う。

テキストマイニングには2つの基本的な特性がある¹⁾。第1に、大量の情報から隠された法則や知見を抽出することである。第2に、このように抽出された法則の中から、意義のある、有意味な、有用な法則を厳選することである。第1の段階では主に統計処理ソフトやアルゴリズムの理解など、数理的な処理(量的分析)が助けとなる。しかし第2の段階では、自明ではないが有意味な法則・解釈・知見は何かという判断が不可欠であり、ここに質的分析も同時に求められていることに留意する必要がある。言い換えれば、テキストマイニングはデータ中心の分析(統計解析)と人間中心の分析(フィールドワークや事例研究等)の中間に位置する「文脈中心の分析」と捉えることもできるだろう(山本 2011, 81)。

現在テキストマイニングは、新聞記事や有価証券報告書、既存のデータベースを対象とした分析等、社会科学でも多くの分野で応用され始めている²⁾。このような潮流は、社会科学でも人文科学でもテキストこそが最も頻繁に研究されてきたデータであること(Wiedemann 2016, 1)を考えれば、驚くべきことではない。ただし、その多くは文化人類学・教育学・社会学・政策科学・経営学の領域に属すものであり、法学・経済学分野における適用例はほとんど見当たらなかった(Ignatow and Mihalcea 2017, 3; Wiedmann 2016, 1)。しかし近年、経済学史の領域でもテキストマイニングを適用した先駆的な研究が現れている。そこで次節以降では、これらの研究を理解しさらに展開していくために不可欠となる、テキストマイニングの諸手法や方法論について順に説明

1) 岡嶋(2006, 28)の指摘による。

2) 日本における社会科学でのテキストマイニングを用いた研究のレビューについては、喜田(2018, 17-28)を参照。同書では①マーケティング、②心理学・福祉・看護・教育、③経済学、④経営学、⑤会計学の各領域で数多くの適用事例が現れていることが示されている。なお経済学史の領域での導入例として、下平・小峯・松山(2012)、古谷(2014)、下平・福田(2014)、小峯・下平(2017)が言及されている。

していく。

II テキストマイニングの諸手法

本節ではテキストマイニングの諸手法を利用（あるいは利用した文献の分析を理解）するために必要な諸概念について、実際にテキスト分析に適用する過程に従って概説する。主な過程としては、①テキストデータを分析可能な形に変換する前処理、②自然言語処理の手法を用いた数量化プロセス（形態素解析、文書単語行列の作成）、③数量化されたデータに対する統計的分析の適用（頻度分析、共起分析等）、④人の手によらないテキストデータの分類・情報抽出（機械学習）がある。なお本節は今後テキストマイニングを経済学史研究者が導入・利用する場合に最低限必要な手法のみに記述を限定しているため、より詳細な分析手法とその解説については、参考文献に記載された専門書等を参照されたい。

1. 前処理

「前処理」はその名称とは異なって、分析の周辺にあるものと捉えるべきなのではなく、精確なデータの収集・分析のために不可欠の部分である。前処理は以下の手続きで進められる³⁾。

- 1) 対象テキストを電子化し、OCR (optical character reader) 処理をする。
- 2) OCR 処理されたテキストを整形（文字化けを直し、無意味な記号を排除するなどの手作業を行うこと）し、テキストファイルとして保存する。
- 3) 保存されたテキストファイルを Microsoft Excel 等の表計算ソフトに読み込ませ、タグ付けを行い「カンマ区切り形式 (CSV)」として保存する。これは汎用性のあるテキストマイニングソフトの多くが、CSV ファイルに対応しているためである。

以上のような前処理を行った後、テキストマイニング用のソフトウェアによって数量化のプロセスが実行される。以下では多くのソフトウェアに共通する分析プロセスと、必要に応じその背景となる理論について説明していく⁴⁾。

2. 形態素解析

文書を数量的に評価し、統計的分析を適用するためには、文書を文字・単語などの単位に変換し数量的な処理ができるようにしなければならない。文を単語ごとに分割し品詞情報などを付け加える作業を形態素解析と呼ぶ（金 2018, 18）⁵⁾。英文の単語を抽出する場合には単語間のスペー

3) 前処理プロセスの課題と手法の詳細については、仲北浦・小峯（2019, 92-96）参照。

4) 日本で利用可能な主なテキストマイニング用ソフトウェアについては、喜田（2018, 18）参照。

5) 形態素解析のプロセスにおいては、検索語として利用すると検索効率が低下するため予め特定の語を除外する処理（ストップワード除去）、同じ語幹・基本形を持つ単語を1つとして数えるための語形変化の解析、などの関連した処理を同時に行う場合がある。

スに基づき分割することができるが、日本語は単語と単語との区切りが明確でないため、日本語に対応した形態素解析用ソフトウェア⁶⁾を使用する。

3. 文書単語行列 (Document-Term Matrix) の生成

形態素解析を行うことにより、文は単語の出現頻度を要素としたベクトルとして表現される。そして分析対象となる文書の各文ベクトルをまとめることにより、全文中に含まれるすべての単語を要素とする「文書単語行列」が生成される。このように分析対象文書を行列として表現することにより、テキストマイニング・ソフトウェアを用いた数量的・統計的な分析が可能となるのである。これは1節で述べた「自然言語処理」と「データマイニング」の切り替え点となる箇所である、ということもできる。以下では、数量的・統計的分析の具体的な内容について説明する。

4. 頻度分析

頻度分析とは、文書内の各単語の出現頻度を計算し、頻度順、時系列順、品詞別等で分類・整理し文書の特性を分析することである。出現頻度の高い単語は重要度が高いと直感的に考えることもでき、文書の特性を分析するための最も基本的な手法であるが、単語間の係り受けや文脈を見ることができず、また低頻度であるが重要である語を見いだせないといった問題がある。

5. 共起分析

頻度分析だけでは文書の特徴が明確にならない場合、単語間の関連性を分析することによって、それぞれの文書の特徴を明らかにすることが可能となる。単語間の共起関係の強弱については、主に2種類の類似度（または距離⁷⁾）によって測定することができる。

- 1) コサイン類似度（距離）：コサイン類似度は、文書単語行列を用い単語間の類似度を単語ベクトルの類似度で数学的に定義する手法である。文書単語行列において各単語は、文書内におけるその単語の出現頻度を要素とするベクトルとして表現される。このとき単語 x を表すベクトル X と単語 y を表すベクトル Y の余弦（コサイン）を考えると、2つのベクトルの向きが近いほどそのなす角度 θ は小さくなる（ $\cos \theta$ は1に近づく）。これにより、以下のようなコサイン類似度を定義することができる（ $\|X\|$ 、 $\|Y\|$ は単語ベクトルの長さ）。

$$\text{コサイン類似度} = (X \cdot Y) / (\|X\| \|Y\|)$$

- 2) Jaccard 類似度（距離）⁸⁾：Jaccard 類似度は、集合概念を用い類似度を測定する方法である。単語 x を含む文書数を X 、単語 y を含む文書数を Y とすると、Jaccard 類似度は以下のように計算される。

6) 現在広く使用されている代表的なものとして、ChaSen と MeCab がある。詳細は金 (2018, 19-20) を参照。

7) 類似度と距離には、「類似度が大きいほど距離が近い」という関係が存在する。

8) 文献や解析ソフトウェアによっては、Jaccard 係数と呼ばれる場合もある。

$$\text{Jaccard 類似度} = |X \cap Y| / |X \cup Y|$$

ここで $|X \cup Y|$ は「[単語 x を含む] か [単語 y を含む] か 一方でも当てはまる文書の数」を表し、 $|X \cap Y|$ は「[単語 x を含む] なおかつ [単語 y を含む] 文書の数」を表す。このように Jaccard 類似度は 2 つの集合に含まれている要素のうち共通要素が占める割合を示しており、これが大きいほど 2 つの単語の類似度が高いと言える。

このように単語間の類似度（距離）を測定しそれにより単語間の共起関係を明らかにすることにより、以下に示すような共起関係に基づく単語や文書の特性（意味⁹⁾ や文脈など）の分析を行うことが可能となる。

6. クラスタ分析

クラスタ分析は、個々のデータ（単語や文など）の類似度を距離として表現し、距離の近いデータ同士をまとめて、分析対象のテキストをグループ化する分析手法である（石田・金編 2012, 11）。距離として表現する類似性には様々な種類があり、データの特性に応じて取捨選択する必要がある。選択した距離の種類によって分析結果が大きく変わることもありうるため、クラスタ分析は、唯一の解を求める最終手段というよりは、データの意味づけを模索するための便宜的方法と捉えた方が良い（石田・小林 2013, 112）。

7. ネットワーク分析

テキストを解析するには、単語間のつながりや共起関係でその使用法や内容を把握することが必要となる。テキストにおける単語の共起パターンは 1 つの重要な情報であるが、その関連性を分析する手法がネットワーク分析であり、視覚的な考察を助けるのがネットワークマップである。

ネットワークマップは文書の中で用いられた単語をノード（結節点）とし、同時に用いられた単語間を結ぶリンク（線）を辺としたグラフである。マップは単語間の共起関係を示す「隣接行列」から得られ、また共起パターンの多少に関する重みを線の太さや長さで示す工夫等も行われている¹⁰⁾。

ネットワークマップで示される構造上の特性を分析するには、主に「中心性」概念と「コミュニティ分析」という手法が用いられる（金 2018, 35-41）。あるノードの「中心性」を測る基準としては、他のノードへの結びつきの多さで測る「次数中心性」、ネットワーク内の情報媒介度で

9) 自然言語処理の領域では、意味は「分布的意味 (distributional semantics)」として捉えられており、ある単語の意味はその周辺に現れる（共起する）単語との関係から表されると考えられている。これはテキストマイニングにおける量的分析が、単語や文書の持つ意味を共起分析を通じて取り扱うことができる根拠となっている（Wiedemann 2016, 25）。

10) ネットワーク分析の理論的基礎と詳細については、金（2018, 34-45）を参照。

測る「媒介中心性」、隣接するノードの中心性を加味する「固有ベクトル中心性」などがある。また「コミュニティ分析」は、ネットワーク内のサブネットワークを検出しそれぞれの特性を分析する手法である。

8. 多変量解析

多変量解析は、複数の変数間に存在する影響・相互作用を一括して解析し、隠された傾向を探る統計的手法の総称である。主成分分析・対応分析・因子分析など数多くの方法がある。いずれの場合も、複数の変数データを何らかの方法で情報の損失を抑えながら少ない変数¹¹⁾に集約したり分解したりすることになる。

ここでは多変量解析の代表例として、テキストマイニングで多く用いられる主成分分析を取り上げる。主成分分析とは、「多くの変数によって記述された量的データについて、複数の変数の相関（共分散）を少数個の合成変数（これを主成分という）に縮約し、データの解釈を用意にするための分析手法」（松村・三浦 2014, 60-61）である。もとの変数が n 個あれば、 n 次元の情報があるといえるが、その情報の大半をごく少数の主成分に要約して表現できれば、この解析は成功したとみなせる。 n 次元をすべて考慮するよりも 2つ3つの主成分のみを熟慮することができ、また散布図¹²⁾などによって全体の傾向を視覚的に表現することが可能となる。

実際の分析においては、文書単語行列の行要素または列要素に主成分分析を適用し、主成分の分布を表または散布図で表すことにより文書の特性を考察する、という手法が使われることが多い。

9. コンコーダンス分析

形態素解析の問題点として、ある語がどのような文脈で用いられているかを検証できないことがあげられる。例えば、個々の単語の意味は、その単語が使われている文脈によって変化する場合が多い。そのため、テキストマイニングを行う際には、頻度という数的に変換された結果を見るだけでなく、可能な限り自分の目で実際の用例を確認する必要がある。

テキストマイニングで用例を確認する場合には、KWIC コンコーダンスという方法が用いられる。「KWIC」は「Key Word In Context」の略であり、「コンコーダンス」は「用例索引」という意味の単語である（小林 2018, 59）。この方法では、注目する単語が PC 等の出力画面の中央に縦に並べられるため、その左右を見比べることで、その単語がどのような文脈で用いられているか

11) この抽出された変数が、多数のデータの中から見いだせる知見となる。その際、各データの平均値だけでなくむしろ偏差（平均値からの離れ具合）を重視することになるが、それはこの偏差が各データの「個性」、「ばらつき具合」を示すからである（涌井・涌井 2011, 19）。

12) 2種類の項目を縦軸と横軸にとり、1つの要素を打点（plot）として記入する図のこと。打点間の相関関係を見いだすことができる。

を簡単に確認することができる。

またこの他にもソフトウェアによっては、KWIC コンコーダンスで指定した単語が用いられている全文を表示する機能や、指定した単語の前後に頻出する語をリストアップする機能(コロケーション統計)、複合語を検出し頻度順に配列する機能(複合語検索)等を備えている場合があり、これらも単語が用いられている文脈を探る場合に有益である。

10. 機械学習 (Machine Learning)

機械学習は、「明示的にプログラムしなくても学習する能力を与えたコンピュータ上のシステム」(高野 2017, 11) と定義され、以下の2つの特徴を持つ処理系としてとらえられる。

- 1) 学習 (Training) : 入力データに基づき、利用側の目的とする事柄に対して、その特徴、時系列や空間上のパターンを学習する。
- 2) 推論 (Inference) : 学習結果に基づき、入力データに対して利用側の目的とする事柄に対する推論を行う。

機械学習の技術を用いることにより、過去の新聞記事や議会資料集等の手作業では扱えないような大量のデータを高速で処理できるようになり、さらにパターンを発見するための十分な量のデータを用意することにより、より高い精度で予測を行うことも可能となる。

機械学習の手法は、教師あり学習 (supervised learning) と教師なし学習 (unsupervised learning) の2つに大きく分けられる¹³⁾。以下では、テキストマイニングにおいて用いられる代表的な機械学習の方法として、教師あり学習によるテキスト分類および教師なし学習であるトピックモデルについて紹介する。

(1) 教師あり学習によるテキスト分類

教師あり学習は、テキストマイニングにおいては特に大規模データの分類手法として用いられる。テキストのカテゴリー (目的変数) が既知であるデータを用いて、カテゴリー別に分類するモデルを作成し、カテゴリーが未知であるテキストを既知のカテゴリーに振り分けることをテキストの分類 (text classification) またはテキストのカテゴリー分類 (text categorization) と呼ぶ (金 2018, 123)。文書だけでなく段落、文、単語もまた分類の対象となり、また分析者の手作業によるカテゴリー分けを補完する役割も果たす。

(2) トピックモデル

トピックモデルは、テキストに含まれるトピックを教師なし学習で推定する方法である (小林 2017, 214)。特に分析データをどのようなカテゴリーに分ければよいか分からない時に有効な手法であり、代表的な手法として潜在ディリクレ配分法 (latent Dirichlet allocation, LDA) がある。LDA ではテキストに複数の潜在的なトピックが含まれていること、また1つのトピックには複数の特徴的な単語が含まれていることを仮定しており、個々の単語の出現頻度や共起パターンを

13) ここで「教師」という用語は、コンピュータがパターンを発見する際の「正解データ」を意味している。

通じて文字列の背後にあるトピックを求めようとする。例えば分析対象の文書に A, B, C という 3 つのトピックと、それぞれのトピックを特徴づける単語が存在している場合、LDA ではそれぞれの文書の特徴づける複数のトピック（文書 1 はトピック A, 文書 2 はトピック B・C…、など）と、それぞれのトピックを特徴づける複数の単語（トピック A は単語 a, b, トピック B は単語 a, c, d …など）の両方を推定し、そこからトピックと文書の分類を行う（金 2018, 84-92）。

以上テキストマイニングで用いられる諸手法を概観したが、次節以降ではこれらの手法を社会科学において用いる際の方法論、およびそれを踏まえた経済学史への適用の方向性について述べていく。

III テキストマイニングの方法論

社会科学におけるテキストマイニングの方法論についてはその導入の初期の段階から様々な議論が展開されてきたが、ここではこれらの議論を踏まえた上で近年体系的な方法論を展開している樋口（2014）、Wiedemann（2016）、喜田（2018）の論を代表的事例として取り上げる。

1. 樋口（2014）による「計量テキスト分析」と「接合アプローチ」

樋口が提案する「計量テキスト分析」の特徴は、テキストマイニングの技術を活用しつつ伝統的な内容分析（content analysis）の考えを実践に生かす点にあるとされる。内容分析とは、文章・音声・映像などのさまざまな質的データを分析するための方法である（樋口 2014, 1）¹⁴⁾。内容分析ではデータをいくつかのカテゴリーに分類したうえで、各カテゴリーのデータの個数を数え上げるなどの計量的分析を行う。現代の内容分析に直接的に寄与した初期の試みは、19 世紀から 20 世紀初頭の新聞記事の計量的分析であり、その後 20 世紀前半に社会科学の理論・概念に加え心理学実験や市場調査のフィールドから優れた統計手法が導入され、その方法論が確立された。現在では内容分析は、アンケートや心理学実験の回答分析、文化人類学、歴史学、文学、実務的分野等多岐にわたる分野で用いられている。

内容分析のうち、一般にテキスト型データのような質的データを分析する場合には、素データの中から分析者が典型的であると考えられる箇所を引用し解釈するという質的な方法を取ることが多かった。これに対し樋口（2014, 5-6）は、計量的分析手法を用いることにより、研究者がたどったデータ収集・分析の過程を他の研究者が検証することが可能となり、これによりその信頼性ないし客観性を向上させ、直接の比較検証に耐える研究を蓄積できることを利点の 1 つとして挙げている¹⁵⁾。

14) 内容分析の定義は時代と共に変化しており、その変遷と評価については、樋口（2014, 3-5）を参照。

15) また計量的分析を導入する場合のもう 1 つの利点として樋口は、データ探索を行えることを挙げている（樋口 2014, 6-7）。

内容分析においては、量的方法と質的方法とが互いに相容れない排他的なものであるとは捉えていない（樋口 2014, 7）。可能な限り厳格な形で量的分析を行ったとしても、研究のさまざまな段階で、量的ではない質的な作業が必要となる。

まず質的なデータを量的に扱える形に変換するという作業は、決して純粋に量的な作業ではないことが指摘される（樋口 2014, 7）。この段階ではいくつかのカテゴリーにデータを分類する「コーディング」という作業が行われるが、この作業を実際に行う前には、いくつかのカテゴリーを作成し「特定の記述がデータ中にあればそのデータを特定のカテゴリーに分類すること」というような基準を作成する必要がある。コーディングの基準を作成するという作業は、研究者の想像力が発揮される、きわめて重要な、なおかつ質的な作業であると考えられる。

また量的分析が完了した後の段階を考える場合でも、やはり量的ではない質的な作業が必要とされる（樋口 2014, 9）。量的分析の結果について解釈を行い結論を導くのは研究者であり、この過程で解釈そのものを生み出す取り組みのほかに、解釈を行うために素データを読み返したり、解釈の妥当性を確認するためにデータに当たるといった質的な作業が必要となる。また量的分析を通じてデータの質的側面に関する新たな発見が得られるのもこの段階であると考えられる。

こうした議論を踏まえ樋口は、内容分析の概念をベースにそこにテキストマイニングに代表される計量的分析手法を導入した「計量テキスト分析」という方法を提案している。計量テキスト分析は、「計量的分析手法を用いてテキスト型データを整理または分析し、内容分析を行う方法である」（樋口 2014, 15）と定義される。この定義では「内容分析を行う方法」という言葉により、計量テキスト分析を内容分析の一種として位置づけるとともに、「計量的分析手法を用いてテキスト型データを整理または分析」という言葉により、量的方法を用いることが明示化されている。

さらに樋口はテキスト型データを分析するための新たなアプローチとして、より客観的な「コーディングルール（分類基準）」に基づく「接合アプローチ」を提唱している（樋口 2014, 17-19）。分析者が作成したコーディングルールに従って言葉や文書を分類するためにコンピュータを用いるアプローチは Dictionary-based アプローチと呼ばれており、ここでは、コーディングルールを作成することで分析者の持つ理論や問題意識を操作化することが目指される。このアプローチの利点には、コーディングルールを作成することで分析者の理論や問題意識を自由に操作化し、テキスト型データのさまざまな側面に自由に焦点を絞れるということがある。一方、意図的ないしは無意識のうちに分析者の理論や仮説に都合の良いコーディングルールばかりが作成されてしまうという危険性もある。

そこで樋口はこの客観性に関わる問題への対応として、テキストマイニングに代表される多変量解析を用いデータを分類・要約する Correlational アプローチを併用することを提唱した（樋口 2014, 18-19）。多変量解析によってデータ全体を要約・提示した上でコーディングルールを公開するという手順を踏めば、データ全体の中からどの部分あるいはどの側面がコーディングルールによって取り上げられたのかを第三者が把握できるようになり、これにより信頼性・客観性を大きく向上させることができることになる。

そしてこれらの議論を踏まえ、樋口氏は以下の2段階からなる「接合アプローチ」を提案している（樋口2014, 19）¹⁶⁾。

段階1 Correlational アプローチにならない、多変量解析を用いることで、分析者の持つ理論や問題意識の影響を極力受けない形で、データを要約・提示する。

段階2 Dictionary-based アプローチにならない、コーディングルールを作成することで、明示的に理論仮説の検証や問題意識の追求を行う。

2. Wiedemann (2016) によるテキスト分析の類型化と「混合メソッド」

ウィーデマンは1960年代からの社会科学領域におけるテキストマイニング¹⁷⁾の方法論をめぐる議論を批判的に検討し、分析対象となる文書における文脈の取り扱いという観点から、その方法について以下の4類型を明らかにしている（Wiedemann 2016, 34-40）。

- 1) コンピュータ支援内容分析 (Computational Content Analysis, CCA) : 大規模な文書コレクションから単語やその頻度を抽出し観測するという量的データのみを取り扱い、文脈についてはほとんど考慮に入っていない。
- 2) コンピュータ支援質的データ分析 (Computer-Assisted Quality Data Analysis, CAQDA) : 小規模の文書の中から選択を行い、それに対し手作業でコーディングを行うCAQDAにより、言語的・意味的文脈を包括的に検討することが可能となった。コーディングを通じた意味の明示的な理解は——テキスト解釈に関する事前の知識を同時に有する——コーディングを行う分析者の認知力を通じて行われる。分析者は（選択された文書の枠内という）局所的な文脈 (local context) の観察を通じて、限られた事例を深く理解することに注力する。
- 3) 計量語彙アプローチ (lexicometric approaches) : 大規模な文書コレクションに対しキーワード抽出や共起分析、主成分分析などの計量語彙アプローチを適用することにより、統計的に検出される言語データにおける顕著な用例などを帰納的に探索することが可能となった。小規模な文書内の局所的な文脈に代わり、文書コレクション全体を観察することにより初めて得られる包括的な文脈 (global context) が導出される。
- 4) 機械学習により導出される文脈は、これまで述べた3つのタイプの間位置するとみなすことができる。機械学習（特に教師あり学習）は、小規模な文書またはその中のより小さな対象（段落、文など）を観測することで得られる局所的な文脈を基礎として行われる。そして学習モデルを用い複数の文書を同時に観測し集計することにより文書コ

16) 樋口氏が開発したテキストマイニング用ソフトウェア KH Coder は、この接合アプローチを実現するためのソフトウェアとして設計されている（インターネット上のサイト (<http://khc.sourceforge.net/>) より入手可能である）。

17) ウィーデマンはコンピュータ支援テキスト分析 (Computer Assisted Text Analysis, CATA) という用語を用いている（Wiedemann 2016, 4）が、これはテキストマイニングとほぼ同義である（金2018, 3）。

レクション全体から得られる情報が認知され、それが包括的な文脈に基づく分析モデルとして具体化される。一方同時にそのモデルから得られた知見は個々の文書に対しフィードバックされるが、それは包括的な文脈を基礎として各文書のカテゴリー分けを行うことを通じて実行される。

このようなテキスト分析の類型化を踏まえ、ウィーデマンは特に計量語彙アプローチと機械学習アプローチの発展に伴い、質的分析と量的分析とを組み合わせた「混合メソッド (mixed methods)」の有効性が高まっていることを指摘している (Wiedemann 2016, 21-22)。質的分析と量的分析との統合は、テキストマイニングの領域における2つの要素の発展により促進されることとなった。1つは大規模なテキストアーカイブが利用可能となったことであり、これにより研究者がその中のさまざまなデータの比較検討を通じて量的分析を伴った質的分析から多くの洞察を得ることが可能となった。もう1つは、計量語彙アプローチや機械学習アプローチによりテキストの意味や文脈を視野に入れた量的分析が展開されることを通じて、分析者による質的分析を通じた解釈と計算機アルゴリズムが導き出す結果との認識論的な差が徐々に狭まっていることである。

このような背景の下、テキスト分析において質的分析と量的分析とは競合するものではなく補完的な技法であると考えられることとなった。質的分析の観点はテキストデータ内のカテゴリーの存在やその構成を理解するために有益であり、また量的分析はそれらのカテゴリーの関連性や多様性などについての情報を与えてくれるのである。

3. 喜田 (2018) による「混合マイニング」

喜田は、本稿の最初に取り上げたように、テキストマイニングを「テキストデータを、言語処理技術を用いて構造化データ・変数に変換し、それをもとに知識発見、仮説発見および仮説検証を行う手法」と定義したうえで、データマイニングとテキストマイニングの関係について、「テキストマイニングが対象とするテキストデータにおいて構造化できるデータであればデータマイニングとして用いることが可能であり、その意味でテキストマイニングはデータマイニングの一部である」(喜田 2018, 300) という立場をとる。このアプローチではデータマイニングの延長線上でテキストマイニングが捉えられており、特に定量的な変数だけでなくテキストを対象にすることも含まれている。そしてこの立場から喜田は数値データとテキストデータを共に用いる「混合マイニング」という方法を提案している。

「混合マイニング」の一連の流れは、以下のように説明される (喜田 2018, 62-65)。

1) データ収集とデータクリーニング

テキストデータ等の「非構造化データ」を定量的なデータとともに収集する。データクリーニングでは、対象となるテキスト特性の把握、長いドキュメントの場合分析単位の設定等を行う。このような手順を経て、対象となるテキストの分析用データを作成する。

2) 概念選択

テキストマイニングにおいては概念選択が重要な問題であり、これを解決する方法を提示することが分析に大きな影響を持つ。このための方法として、「データビジュアライゼーション(データオリエンテド)」と「セオリーオリエンテド」という2つの方法が提案されている。前者の方法は言及頻度による概念選択であり、言及頻度の上位順に概念を選択し、それをもとにテキストマイニングを行う方法である。後者は、先行研究や背景知識などによって概念選択・分析視点を理論的に構築する方法である。

3) 非構造化データの構造化

非構造化データであるテキストデータを構造化する段階であり、形態素解析による定量的な変数への変換が行われ、文書単語行列等によって表現される構造化データが作成される。このようにして統計的分析が可能となれば、それに基づく分析・検証を行うことになる。

4) モデリング手法とデータビジュアライゼーション

構造化データによる統計的分析のほか、機械学習などのアルゴリズムを用いて予測・判別、話題の分類、パターン発見などの手法も適用できる。また分析結果を報告や論文にする場合、グラフ化セオリー等を用いた分析結果の見せ方を検討する必要がある。

以上近年におけるテキストマイニングに関する方法論を扱った事例を取り上げてきたが、特に3者に共通しているのは、テキストマイニングの諸手法の発展を踏まえた質的分析と量的分析との接合(あるいは混合)を強調していることであり、このアプローチが今後社会科学における主流となる可能性を示唆していると捉えられる。この中でも特に樋口(2014)の「計量テキスト分析」と「接合アプローチ」は、「理論仮説の検証や問題意識の追求」が要求される経済学史研究に親和的であると考えられ、今後の一層の発展と応用が期待される。

IV 経済学史研究へのテキストマイニングの適用

ここまでテキストマイニングの手法と方法論について概観してきたが、本節¹⁸⁾では、III節で述べた質的分析と量的分析との接合という方向性のうち特に樋口氏の「接合アプローチ」の応用を視野に入れた、経済学史研究にテキストマイニングを適用する意義と方法について述べる。次に近年経済学史研究においてテキストマイニングを導入した分析事例を紹介し、具体的な導入手法や分析プロセスを明らかにしたい。

1. 経済学史研究へのテキストマイニング適用の意義と方法

経済学史の研究者は、伝統的に地道なテキスト再構成に携わってきた。経済学史家による解釈(再構成)が公表された時、通常科学と同じく、批判と競争を通じてより良い認識が徐々に積み

18) 本節1項は下平・小峯・松山(2012, 6-10)、小峯・下平(2017, 5-18)に加筆・修正を加えたものである。

上がってきた場合も多いだろう。ただしこの再構成を実行するのは、いずれの場合も現在の研究者である。再構成に当たって、我々自身の通念・世界観が反映され、時代の制約を受ける。どんなに広い視野や自省や批判的視座を持っていても、逃れられない盲点は存在する。

この盲点を解消させる1つの手段として、テキストマイニングの適用を考えてみたい。定性的(質的)分析と分類される通常の経済学史研究を進めた後、一旦は個人の意思とは離れた量的分析をテキスト再構成の過程で噛み合わせるならば、全体としての解釈に合理性・一貫性・説得性を増加させることが可能となるのではないか。例えば、我々が重要とはみなさない「過去の専門用語・日常用語」を通常のテキスト解釈では見逃しても、機械的に処理されたテキスト分析によって「新たなデータとして発見」することが可能になるかもしれない。ここにテキストマイニングという技法を採用し、従来の経済学史研究と接合させる可能性が見いだせる。

このようなテキストマイニングの手法は、従来からの経済学史研究にどのような意味を持つのだろうか。結論を先に言えば、二重の相互作用¹⁹⁾によって、望ましい相乗効果が期待できるだろう、という意義を持つ。

第1の相互作用は、質的分析と量的分析との間に発生する。テキストマイニングの方法ではデジタル化されたデータ収集から始め、統計的手法を導入するための前処理を行い、そのデータを解析して結果を得る。特に単語レベルの関係性が瞬時に正確に抽出される²⁰⁾。この過程で、各テキストは形態素に分解され、文脈レベルでも単語レベルでも切り離され、一旦意味が喪失している(「二重の喪失」²¹⁾)。つまり、テキストマイニングによる自動的な解析は、それ自体ではあまり意味を持たない。実際には、テキスト全体に対して、様々な段階で「解釈」を施さなければならぬ。すなわち、仮説、収集されたデータそのものの意義、カテゴリーの名前付け、結果の意義と限界、そしてテキスト全体の総合的解析という各場面で、解釈が重要な位置を占める。

換言すれば、一旦解体されたテキストを再構成²²⁾するためには、解析者の観点・目的が必要不可欠なのである。その文章の意義を予め熟知している解析者は、テキストマイニングの前に「柔軟な作業仮説」を既に持っている。その観点から、次の(グループ分けや新しい文書との比較等を含めた)分析方法が定まり、十分に精確なテキストマイニングが可能になり、「より固まった結論」を導ける可能性が高まる²³⁾。質的分析と量的分析とは、言わばサンドイッチのように挟

19) 「二重の相互作用」の図式的説明については、小峯・下平(2017, 15)を参照。

20) 佐藤(2008, 57)は、テキストマイニングの特徴(限界)として、基本的な分析ユニットが単語(テキストの最小単位)であることを指摘している。

21) この表現は菰田・那須川編(2014, 16)から示唆されたものである。

22) 佐藤(2008, 50)は、この過程を文書セグメントの「再文脈化」と呼び、第1段階(データベース化; 分類と配列)と第2段階(ストーリー化; 新しい意味づけ)に分けた。

23) 佐藤(2008, 12)は優れた質的研究のためには、手持ちのデータで全体の筋を洗いデッサンのように一旦仕上げてみて問題点を浮き彫りにさせた後、データの追加的収集と解析に進んでいくべき、と述べている。このような相互作用のため、解析技術を熟知したものとテキストの意義を熟知した者が分業すれば効率的であろう。

み込まれた関係が必要となる。この意味で、解析者である研究者の内部に、質的分析と量的分析との相互作用・相互交流を意図的に発生させなければならない。

第2の相互作用は、研究者個人を媒介として、原典と研究者共同体との間に発生する。経済学史の研究は、過去の経済学者による原典が持っている個別の意味・意義・価値観・世界観・社会的重要性を、研究者個人の観点や分析を通じて、現在の学問という研究者の共同体に共有される言葉、つまり普遍的で抽象的な知・知見に変換していくことであり、同時にそれを研究者共同体の先に、我々自身が生きている社会にその知見を届けることでもある。この変換には必ず、研究者共同体のロジック・世界観からの方向性もある。2つの世界を、研究者個人とその言葉を媒介にして何度となく往復できる場合、「厚みのある記述」が可能になるのではない。

このようにテキストマイニングという新しい技術と従来の経済学の知見を交互に組み合わせ、相乗効果を発揮させれば、従来は見逃されていた新しい知見——例えば、匿名文書の筆者の推定、当該書籍の版別異同における思想的変化の明確化、通俗化・単純化された支配的言説の変形など——が生まれる可能性を秘めている。つまり、研究者の直観や長年の蓄積によって支えられてきた質的なテキスト解釈の途中に、大量のデータを統一的に処理するという量的分析を挟み込むことによって、より深みのある、説得力の増す仮説・知見を見いだすことができるのではないか。さらに、この知見が原典と研究者共同体をさらに緊密に結びつけ、通常は顧みられることの少ない原典の現代における意義を再確認できるのではないだろうか。

この意味で、テキストマイニングは経済学史研究と適合的であり、より「発見的 (heuristic)」な知見を促す新しい手法となりえよう²⁴⁾。そして、研究手法においても多様なアプローチがあることを示せば、過去の経済学者たちの主張を、より客観的・効果的に掘り起こす手法が競合する余地を生じさせ、様々な分野の研究者が経済学史・経済思想史という分野に参入する契機となりうるだろう。留意すべきは、この質的分析と量的分析とは背反ではなく、むしろ相乗効果を持つ点である。テキストマイニングという新しい手法は、従来から蓄積してきた質的分析と結合することで、経済学史という分野自体の伝播を広げる可能性を持つだろう。

そしてこれまで述べてきたテキストマイニングを経済学史に導入する意義を視野に入れ、仲北浦・小峯 (2019, 7-8) では、樋口氏の「接合アプローチ」を経済学史研究に統合する4段階からなる分析方法を提案している。まず第1段階は「柔らかな作業仮説」であり、当該テーマに関する先行研究をふまえて、分析者自身の問題意識に基づいた仮説を明らかにする。続いて第2段階は「機械的な量的分析」、第3段階は「質的カテゴリーを加えた量的分析」を行うが、この段階で樋口氏の「接合アプローチ」が適用される。そして最後に第4段階として、第2段階と第3段階の解析結果を踏まえて原典の内容分析を行い、分析者の主張を根拠づける「原典による引証」

24) 村井編 (2014, 140) も同じ立場にあるが、あくまでもこの方法は期待される効能であり、逆に、テキストマイニングという手法が、従来の経済学史研究と比べて有効でないと判断する場合もありうる。この場合、旧来の計量文献学に代表されるように、匿名の文書・パンフレットの著者特定などという限定された領域において、この手法を経済学史研究に適用することになるだろう。

を行う。

このような分析を通じて、従来の経済学史研究の方法とテキストマイニングの方法を結びつけることができると同時に、4つの段階を踏むことにより分析者の主張がより説得的になるという意味で、従来の経済学史研究の方法をより強化できる可能性があるだろう。

2. 経済学史研究におけるテキストマイニングの導入事例

(1) 経済学の大衆化に関する分析——下平・福田 (2014)——

下平・福田 (2014) の課題は、テキストマイニングを用いて古典派経済学の普及過程を検討することであり、特にリカードウの経済学が通俗化していく過程の一端を明らかにすることであった²⁵⁾。そのために、テキストマイニングを用いて、リカードウ『経済学および課税の原理』(第3版)(以下『原理』と略記)、ジェイムズ・ミル『経済学要綱』(第3版)(以下『要綱』)、ハリエット・マーティノー『経済学例解』(第3版)第25話(以下『例解』第25話)についての比較検討を行った。マーティノーは古典派経済学の普及のためにとくに大きな役割を果たした人物として知られているが、リカードウ『原理』の議論がマーティノー『例解』によって広められる際にどのような変容を被ったのか、こうした過程をミル『要綱』を間に挟んで明らかにすることを意図している。

次に設定された課題に基づき、実際の分析を行うに当たり次のような仮説を予め構築した。すなわち、マーティノー『例解』は一般大衆を読者として想定していたため、リカードウ『原理』やミル『要綱』と比較して、

- 1) 専門的な用語を排除し、一般大衆が理解しやすい通俗的な言葉を多く用いている。
- 2) 基礎的な理論分析よりも、その背景にある世界観や政策論に注目している。

その上でこうした仮設が成立するか否か、そしてそれらがテキストマイニングによって正当化されるか否かを探っていこうとしている。具体的には、対象となる文献に対し頻度分析とネットワーク分析を適用し、それぞれの文献の特徴を明らかにしながら比較検討を行った。

テキストマイニング分析とその結果に基づく考察より、リカードウ『原理』、ミル『要綱』、マーティノー『例解』第25話の間の議論の異同が鮮明になった。まず頻度分析により、リカードウ『原理』で出現頻度の高い名詞 (price, value, profit 等) がミル『要綱』ではその順位が分散化し、マーティノー『例解』第25話では大きく順位を下げ、代わって多くの名詞 (society, class, capitalist 等) が新たに上位に入っていることが明らかとなった。次第に順位を下げた名詞の多くは価格決定や所得分配の問題に関わる理論的概念に相当するものであり、順位を上げた名詞の多くは比較的大雑把に社会感を表す現実的概念に相当するものだった。

次に共起ネットワーク分析より、リカードウ『原理』で中心的な役割を果たす名詞 (price, profit) がミル『要綱』では別の名詞 (labor, commodity) に入れ替わり、マーティノー『例解』

25) 研究の背景の詳細については、下平・福田 (2014, 51) 参照。

第25話ではさらに別の名詞 (capital, society) に入れ替わっていることが分かった。ここでも価格決定と所得分配に関わる理論的概念に相当する名詞が、次第に比較的大雑把な現実的概念にその役割を奪われていったと言える。また、リカードウ『原理』の関連名詞のグループ分け (money—value—corn/land—produce—rent 等) が、ミル『要綱』ではある程度引き継がれながらも変化・分散し、マーティノー『例解』第25話では決定的に変化・分散し、全く別のグループ分け (labourer—capitalist—wages/food—supply—demand 等) が形成されていることも分かった。ここでも厳密な理論的概念によって構成された問題群が、次第に分散化され通俗化されていったと言えるだろう。

こうした考察の一定の結論として、リカードウ『原理』の専門的な議論は、ミル『要綱』において簡略化・平易化され、マーティノー『例解』第25話において決定的に通俗化されることを通して、広く一般大衆に受け入れられていったと言えるだろう。

(2) 『国富論』へのトピックモデルの適用——Binder & Jennings (2016)——

バインダー&ジェニングスは、アダム・スミス『国富論』第3版(1784)の索引と、同書にトピックモデルを適用した場合の分類結果とを比較することにより、テキストマイニングの経済学史研究における有効性を検討している²⁶⁾。索引とトピックモデルの分類とを比較した結果両者が重なる部分も多かったが、同論文では共通点よりもその差異が両者の基本的な相違を明らかにしていると考えている。

『国富論』の索引とトピックモデルの分類との差異に注目した場合、以下のような特性があることが明らかになった。①索引においては1つの概念としてまとめられている事項が、トピックモデルでは別々のトピックとして表現される場合があること、②索引で用いられている文章における比喩や反例について、トピックモデルは正確に分類することはできない場合がある、などである。①については、「労働」に関する索引で同時に取り上げられている分業の具体的説明(ピン製造業の例)とそこから導かれる労働に関する抽象的議論が、トピックモデルでは別トピックとして分類されていること、②については、索引におけるフランス重農主義に関する批判の項目を、トピックモデルでは経済分析と捉えている、という事例で説明している。

ただしこのことは決してトピックモデルの有効性を否定するものではなく、索引とトピックモデルの分類結果を比較可能な形で提示することを通じて、単にソフトウェアが見いだしたパターンだけに依拠せず、2つの見方の間にある相関と矛盾を識別することにより、異なる角度からの解釈比較分析を実行できる可能性を示唆している。またこの比較アプローチは、テキストマイニングに馴染みのない研究者にとってそれを理解するための歴史的展望を提供するためにも役立つとされている。

26) 分析結果については、<http://networkedcorpus.com/networkedsmith/> において閲覧することができる。

(3) 経済学者の言説に対する「接合アプローチ」の適用——仲北浦・小峯 (2019)——

仲北浦・小峯 (2019) の目的は、ケンブリッジ学派の経済学者 D. H. ロバートソンが 1929 年の「金融および産業に関する委員会」(通称「マクミラン委員会」) に提出した証言資料をテキストマイニング分析することで、本委員会の議論における彼の言説 (特にその経済観) の特徴を探ることである。この論文は、先に述べた樋口 (2014) の「接合アプローチ」を基にした 4 段階の分析手法を経済学史研究に適用した最初の事例である。

ロバートソンの理論・思想に関する先行研究において、ロバートソンは実物としての人間の意思決定とそれに基づく活動が経済の根本であると捉えていたという主張がなされており、このことを証言資料を通じて論証するために、第 1 段階として以下の「柔らかな作業仮説」を設定した。

- 1) ロバートソンは、金融の議論においても、実物理論を重視した。
- 2) ロバートソンは、経済現象の説明において、経済主体の意思決定とそれに基づく経済活動を重視した。

このような仮説を念頭に置き、次に『金融および産業に関する委員会報告書』(以下『報告書』と省略) とロバートソンの証言資料「世界不況」のテキストデータに対し多変量解析 (頻度分析、階層的クラスター分析、共起ネットワーク分析、対応分析) を行った。『報告書』を同時に分析対象とする理由は、『報告書』と「世界不況」との論調の差を比較することにより、後者の特徴をより明確にしようとする意図による。この第 2 段階では、分析者の意図をできるだけ混入させずに分析を行っており (機械的な量的分析)、解析の結果、ロバートソンの「世界不況」では実体経済に関する語 (consumption, goods, production, things) が特徴的であることが明らかになった。これにより、ロバートソンは金融問題に関する証言においても財市場を併せて考慮していたことが予想される。特に対応分析において、上記の実体経済に関する語のほか、人間 (経済主体) を表す語 (agency, entrepreneur, labour) および経済生活に関する語 (desire, living, wage) が「世界不況」の特徴語として現れていることが明らかとなった。このことからロバートソンが実体経済のみならず、それを形成する人間 (経済主体) とその活動や経済生活をも議論の俎上にのせている可能性が明らかにされた。

そしてこの結果を踏まえて、第 3 段階では (1) 実物理論、(2) 経済主体、(3) 動機・活動に関する語をまとめた質的なカテゴリーを作成した (コーディングルールを作成)。そしてこのコーディングルールを『報告書』、「世界不況」に適用し、集計と視覚的な分析 (バブル図、折れ線グラフ) を行った。その結果いずれのコードにおいてもロバートソンの「世界不況」における出現割合が最も大きく、作成したコードがロバートソンの特徴とみなすことができた。

そして最終的な第 4 段階として仮説の対象となる文献を精査し、これにより、マクミラン委員会におけるロバートソンの言説の特徴として、金融問題においても実物理論を重視したこと、経済主体としての人間の意思決定とそれに基づく活動を強調したことを確認している。

(4) その他の分析

これまで述べた事例以外にも、テキストマイニングの手法を一部活用し、数量化されたデータを用いた経済学史に関する論文が海外で発表されている。まず Wright (2016) はネットワーク分析を社会的関係に適用し視覚的に表現する「社会ネットワーク分析」を用いて、1920年代のウィーンを舞台に非公式なセミナーに出席した研究者のリストを数量化し、多様な個人と集団が学問分野を超えて交流する様子を描いている。

また Claveau & Gingras (2016) は、計量書誌学と動的ネットワーク分析を組み合わせて、1956年から2014年までの期間で、経済学の内部でどのような専門分野が盛衰するかを描いている。具体的には、諸論文の索引データを基に論文ごとの類似性を計測し、類似した論文のクラスタの時系列的变化を検討することによりそれらの動きを追跡している。

V おわりに

本稿は「テキストマイニング」という分析手法を経済学史研究へ適用することの意義を明らかにし、また実際の分析事例を示してきた。テキストマイニングの方法論に関する先行研究を踏まえ、テキストマイニングを経済学史研究に適用する意義として「定性的（質的）分析と分類される通常の経済学史研究を進めた後、一旦は個人の意思とは離れた量的分析をテキスト再構成の過程で噛み合わせるならば、全体としての解釈に合理性・一貫性・説得性を増加させることが可能となる」ということを明らかにした。またこのような観点から、経済学研究へテキストマイニングを適用する際の「柔らかな作業仮説」、「機械的な量的分析」、「質的カテゴリーを加えた量的分析」、「原典による引証」という4段階の分析手法を提示した。

これまで見たように、テキストマイニングには2つの基本的な特性がある。(1) 大量の情報から隠された法則や知見を抽出すること、(2) このように抽出された法則の中から、意義のある、有意味な、有用な法則を厳選することである。第1の段階では数理的な処理（量的分析）が助けとなるが、第2の段階では自明ではないが有意味な法則・解釈・知見は何かという判断が必要であり、ここに質的分析も同時に求められることとなる。すなわち、量的分析の過程で一旦解体されたテキストを再構成するためには、解析者の観点・目的が必要不可欠なのである。したがって解析者は、テキストマイニングの前に「柔らかな作業仮説」を持ち、その観点から分析方法を定めなければならない。それにより十分に精確なテキストマイニングが可能になり、「より固まった結論」を導ける可能性が高まるのである。

このようにして、研究者の直観や長年の蓄積によって支えられてきた質的なテキスト解釈の途中に大量のデータを統一的に処理するという量的分析を挟み込むことによって、より深みのある、説得力の増す仮説・知見を見いだすことができるのではないか。さらに、この知見が原典と研究者共同体をさらに緊密に結びつけ、通常は顧みられることの少ない原典の現代における意義を再確認できるのではないだろうか。このような観点から、テキストマイニングは経済学史研究と適

合的であり、より発見的な知見を促す新しい手法となりうる、と現時点では結論できるだろう。

今回見たように経済学史研究におけるテキストマイニングの導入事例はまだ少なく、また現状では、① テキストの電子化という初期条件がまだ十分に満たされていない、② 対象文献の権利関係の取り扱いが明確ではない²⁷⁾、③ 新たな知識や視点の発見まで至る事例が少ない、といった課題も存在する。今後より多くの研究者がテキストマイニングを経済学史研究に導入し、それらが批判的に検討されることを通じて新たな展望が開けることを期待したい。

(下平裕之：山形大学)

参 考 文 献

- Binder, J. M. and C. Jennings. 2016. "A Scientific View of the Whole": Adam Smith, Indexing, and Technologies of Abstraction. *Journal of English Literary History* 83 (1): 157-80.
- Claveau, F. and Y. Gingras. 2016. Macrodynamics of Economics: A Bibliometric History. *History of Political Economy* 48 (4): 551-92.
- Ignatow, G. and R. Mihalcea. 2017. *Text Mining: A Guidebook for the Social Sciences*. California: SAGE Publications.
- Matsuyama, N. 2016. A Study of Text Mining for Research into the History of Economic Thought: The Case of Alfred Marshall's *Principles of Economics* (1890). *Discussion Paper*. University of Hyogo, No. 89.
- Shimodaira, H. and S. Fukuda. 2014. Popularization of Classical Economics: A Text-mining Analysis on David Ricardo, James Mill, and Harriet Martineau. *Discussion Paper Series*, Research Group of Economics and Management, Yamagata University, No. 2014-E01.
- Wiedemann, G. 2016. *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*. Wiesbaden: Springer.
- Wright, C. 2016. The 1920s Viennese Intellectual Community as a Center for Ideas Exchange: A Network Analysis. *History of Political Economy* 48 (4): 593-634.
- 石田基広. 2008. 『Rによるテキストマイニング入門』森北出版.
- 石田基広・金 明哲 編. 2012. 『コーパスとテキストマイニング』共立出版.
- 石田基広・小林雄一郎. 2013. 『Rで学ぶ日本語テキストマイニング』ひつじ書房.
- 岡嶋裕史. 2006. 『数式を使わないデータマイニング入門—隠れた法則を発見する』光文社新書.
- 喜田昌樹. 2008. 『テキストマイニング入門』白桃書房.
- . 2018. 『新テキストマイニング入門—経営研究での「非構造化データ」の扱い方』白桃書房.
- 金 明哲. 2009. 『テキストデータの統計科学入門』岩波書店.
- . 2018. 『テキストアナリティクス』共立出版.
- 小林雄一郎. 2017. 『Rによるやさしいテキストマイニング [機械学習編]』オーム社.
- . 2018. 『Rによるやさしいテキストマイニング [活用事例編]』オーム社.
- 小峯 敦. 2014. 『『ベヴァリッジ報告』(1942)と『雇用政策』白書(1944)—戦後構想(社会保障と完全雇用)における経済助言活動の役割』『龍谷大学経済学論集』53 (1/2): 37-98.
- 小峯 敦・下平裕之. 2017. 「ベヴァリッジ『自由社会における完全雇用』のケインズの要素—テキストマイニングを加味した量的・質的分析」龍谷大学経済学部 Discussion Paper Series, No. 17-01.
- 菰田文男・那須川哲哉 編. 2014. 『ビッグデータを活かす技術戦略としてのテキストマイニング』中央経済社.

27) この課題については、仲北浦・小峯(2019, 43-45)において検討が行われている。

- 佐藤郁哉. 2008.『質的データ分析法』新陽社.
- 下平裕之・小峯 敦・松山直樹. 2012.「経済学史研究におけるテキストマイニング分析の導入—ケインズ『一般理論』と書評の関係」Discussion Paper Series, Research Group of Economics and Management, Yamagata University, No. 2012-E02.
- 下平裕之・福田進治. 2014.「古典派経済学の普及過程に関するテキストマイニング分析—リカード, ミル, マーティノーを中心に」『人文社会論叢 社会科学篇 (弘前大学人文学部)』31:51-66.
- 高野茂幸. 2017.『Thinking Machines—機械学習とそのハードウェア実装』インプレス R&D.
- 仲北浦淳基・小峯 敦. 2019.「マクミラン委員会におけるロバートソンの経済観—テキストマイニングから判明する「実物」と「貨幣」」『経済学論集』(龍谷大学経済学部) 58 (1): 59-100.
- 樋口耕一. 2014.『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して』ナカニシヤ出版.
- 古谷 豊. 2014.「テキストマイニングを用いたスミス『国富論』普及の分析」TERG Discussion Papers No. 325.
- 松村真宏・三浦麻子. 2014.『人文・社会科学のためのテキストマイニング [改定新版]』誠信書房.
- 村井 源 編. 2014.『量から質に迫る—人間の複雑な感性をいかに「計る」か』新曜社.
- 山本真照. 2011.「テキストマイニング手法の洗練に向けた知識活用方法に関する研究」『経済科学論究』(埼玉大学経済学会) 8:73-85.
- 涌井良幸・涌井貞美. 2011.『多変量解析がわかる』技術評論社.