

テキストマイニングによる文献研究

— 営業研究分野への適用事例 —

北中英明

要 約

文献研究（法）は、研究活動を遂行する際の最も基本的な研究活動である。それにもかかわらず、文献研究の方法論については、各研究者が経験を通じて体得するものだと認識され、用いられるアプローチについても、あまり顧慮されることがない。本研究は、文献研究法に機械学習の一種であるテキストマイニング手法を適用し、その潜在的な可能性と有用性について考察した。テキストマイニングでは、確率的潜在意味解析（潜在ディリクリ配分法：LDA）という手法を用い、営業研究に特化した学術専門誌の掲載論文を対象に分析をおこなった。分析の結果、研究動向を指し示す五つのテーマを抽出することができた。文献研究にテキストマイニングを用いることの有用性は高く、今後盛んになる可能性がある。

キーワード：文献研究，営業研究，テキストマイニング，確率的潜在意味解析，潜在ディリクリ配分法（LDA）

1 文献研究法について

文献研究とは、おもに学術雑誌を対象に、過去の学術論文を系統だって調べる行為である。文献研究は文献研究、文献サーベイもしくは文献レビューなどと呼ぶ。こうした作業によって執筆された論文は、サーベイ論文（survey article）、レビュー論文（review article）などと呼ばれる。レビュー論文（review article）といった場合には総説論文的な意味合いの場合が多い。展望論文という場合もある。文献研究法¹⁾は、文献研究を用いる研究手法のことを指し、多くの学術領域において実践されている。

文献研究（法）は、研究活動を遂行する際の最も基本的な研究活動である。特に、研究論文を執筆する上では欠かせない、重要な研究活動である。そのため、あらゆる研究領域において実践されている。その反面、文献研究の方法論については各研究者が経験を通じて体得するものだと認識されている。また、文献研究に用いられるアプローチについても、従来の伝統的なアプローチが使わ

1) 文献研究に似たものに文献学がある。文献学とは、文献の原典批判・解釈・成立史・出典研究をおこなう学問である。

れ、それほど顧みられないことがない。

1-1 文献研究の種類と目的

文献研究論文は、以下の3種類に分類される。

① 総説、レビュー

総説論文は、当該研究分野の権威と目される研究者によって書かれる場合が多い。単体のジャーナルを対象として分析する場合と、複数のジャーナルを横断的に分析する場合がある。過去の研究動向やこれからの研究の方向性、残された研究課題をまとめた内容となっている。学会、専門誌の研究動向をまとめて記録に残すことの意義は大きい。先行研究から得られた知見を、わかりやすい形で蓄積できるため、学会や当該研究分野の発展に寄与することができる。また、啓蒙といった面もあり、当該分野の初学者にとっては有益である。

② 学術論文の一部

通常の学術論文の場合、先行研究と呼ばれる文献研究をおこなう。先行研究がない論文は、学術論文であるとはいえない。先行研究によって自分の当該研究についての知見、課題を得ることが重要である。

先行研究をおこなう目的は、当該分野の研究動向を知ること、当該分野における残された研究課題を見つけること、研究手法についてのアイデアやヒントを得ること、などがある。また、研究分野によっては過去の研究の再試験・追試験をおこなうために先行研究をおこなう場合もある。

先行研究についての概観をまとめて、通常論文の前半部分におかれる。学術論文の基本的な構成はIMRAD²⁾といわれている。一般に学術論文の一部としての(先行)文献研究は、Introduction(1)に置かれる。

③ 独立した研究としての文献研究法

文献研究を、定量的研究、定性的研究に並んで第三の研究手法という意見もある(大木, 2013)。これは上記①と似ているが、①の場合はあくまでも、対象研究分野の動向を把握するのが目的であり、それを踏まえて別途研究をおこなうという態度であるのに対して、この考え方は、文献研究そのものが一つの研究業績として成立するという考え方である。

1-2 文献研究の方法論

文献研究をおこなう際に重要なことは、次の通りである。

- ① 網羅的に収集する
- ② 批判する
- ③ 研究領域での全体的な構図を作る

2) それぞれ Introduction, Methods, Results, And Discussion の頭文字を取った略語である。

④ 隙間、漏れている箇所を特定し、リサーチクエスチョンにつなげる

特に重要な考え方として、システマティック・レビュー (systematic review) という考え方がある。無造作に、関連する文献を探して読むのではなく、対象領域・対象テーマに関して系統的に渉猟する方法論であるといえる。これは、伝統的なレビュー (traditional review) や物語風のレビュー (narrative review) とは異なる。研究手法として文献研究を用いる際には必須である。また、メタアナリシスも重要であるとされる。メタアナリシスとは、複数の研究の結果を統合し、より高い見地から分析すること、またはそのための手法や統計解析のことである。システマティック・レビューとの関連では、メタアナリシスは特にデータ分析をすることを指して使う場合が多い。

1-3 質的研究と量的研究

近年になって、多くの学問領域で科学性が問われている。社会科学 (経営学・商学) においても同様である。科学性に関しては、量的研究の方が有利であり、今日では量的研究の方が多勢を占めている。一方、質的研究の陣営からは、研究方法論としての足場を築き正当性を獲得するための活発な主張がなされている (フリック, 2011; プレイディー&コリアー, 2008; キング他, 2004)。量的研究と質的研究の主な違いは、次のようにまとめられる。

ここで、量的研究と定性的研究についてのそれぞれの特徴を比較しながら簡単にまとめておきたい。まず、量的・定性的という用語であるが、一般的に前者は伝統的で、実証主義、実験主義、経験主義的なアプローチに用いられ、後者は、構成主義的なアプローチで、自然主義、解釈主義、ポスト実証主義、ポストモダン主義といった見方に対して用いられる。量的研究の具体的な手法としては、①実験と②アンケート調査やインタビューがある。定性的研究の主な研究手法としては、①理論構築、②文化人類学的アプローチ、③現象学的アプローチ、④アクション・リサーチ、⑤解釈学、⑥ケース・スタディー、といったようなものがある。(北中, 1998)

また、量的研究と質的研究の背景にある考え方の違いは次のようにまとめられる。

存在論的な考え方については、量的が現実を客観的なものであると見なすのに対して、定性的研究は現実を個々の研究に携わる研究者のみによって作られ、その結果複数の現実が存在するとする。研究者と研究対象の関係については、量的研究では、研究者は研究対象から独立しておりしたがって客観的な存在であるのに対して、定性的研究では、研究者と研究対象は相互に作用しあうと考える。これは価値についての態度にもつながっており、量的研究では研究者の価値は遠ざけられる (value-free) のに対して、定性的研究の性質は価値負荷的 (value-laden) なものであるとする。定性的な研究は、現実とは社会的に構成されたものであり、調査者と調査対象との良好な関係に重点を置く。その結果、調

査者は価値負荷的なものを見方を強調するのである。(北中, 1998)

双方の研究手法に対しては、互いに批判的・否定的に観る場合が多い。そうした批判を集約すると、質的研究に対する批判は、特に多数を占める事例研究に対して、少ない事例による知見の一般化や事例の選択基準についての合意が得られない、などであろう。量的研究に対する批判は、分析の背景に仮定としておいている統計的前提である。例えば、研究対象の事例が本当に正規分布に従うのか、等である。なお、研究手法の選択については、明確な基準や正解がないとされている。どちらの研究手法のも一理あり、結局は好みの違いという側面もある。

質的研究法と量的研究法の両方をおこなう混合研究法 (Mixed Methods Research) が提案されている。文献研究に関しては、一般的に質的研究の中にも含む場合が多いが、混合研究法によると、文献研究のような文書の内容分析研究は混合研究法の一例であると分類されている (クレスウェル & クラーク, 2010)。本研究では、テキストマイニングによって文書 (定性データ) を分析する。テキストマイニングは、機械学習という定量的なアプローチの一つである。従って本研究の場合は、まさに混合研究法そのものといえるであろう。さらに、本研究のアプローチは確率的潜在的意味解析 (後述) であり、その意味では研究者の価値が遠ざけられた (value-free) なアプローチであると言える。こうした人間の手を介さない手法によるアプローチは、分析に際して研究者の主観や好みによるバイアスを避ける心乙が期待できるため、有益であると考えられる。

2 テキストマイニングの概要

2-1 テキストマイニングとは

今日はビッグデータの時代である。多種多様なデータが大量にかつ容易に入手出来るようになってきている。WWW や企業内の情報共有掲示板 (かつてはイントラネットと呼ばれていた) 上に大量のデータが存在している。特にブログの普及や SNS の利用者拡大によって、多くの人が書き込むコメントが大量に生成され蓄積されている。さらに電子配信されるニュースサイトの記事などが大量に生み出されている。こうしたデータに加えて、従来から企業や役所等の組織では、さまざまな内容の出来事の記録が文書として大量に生成され保管されている。これらの近年になって新しく生み出された大量のデータや従来からの文書がテキストデータである。テキストデータを一連の系統だった手続によって分析しようとするのを、テキストマイニングと呼ぶ。

テキストマイニングは、文書をコンピュータを用いて分析するための手法である。テキストマイニングの手法は、クラスタリング、コレスポネンス分析、自己組織化マップ、共起ネットワークなどがあり、近年では潜在的意味解析に統計的 / 確率的モデルを用いたものなどさまざまである。また、テキストマイニングの応用分野は、文書の分類にとどまらず、消費者の購買履歴や SNS 上の話題の変遷等、多岐にわたる。そのため、ビジネス界からも、その応用に対する期待が高まっている。また、オープンソースの Python の普及により、高価な市販ソフトウェアを購入する必要性が減少している。そのため、多くの人がテキストマイニングに手軽に取り組むことが可能になっており、それが利用拡大を促進している。

テキストマイニングは、以下の様に定義される³⁾。

テキストマイニングは、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であり、自然言語処理、統計解析、データマイニングなどの基盤技術の上に成り立っている。(松村・三浦, 2014, p. 1)

もう少し詳しい定義では、下記のようになる。

テキストマイニングは、この情報爆発の問題を解決する新しい魅力的な研究分野であり、そこではデータマイニング、機械学習、自然言語処理、情報探索、ナレッジマネジメントといった分野で培われてきた技術が用いられている。テキストマイニングは、文書集合の前処理(テキスト分類、情報抽出、用語抽出)、中間表現の保管、中間表現の分析(分散分析、クラスタリング、傾向分析、相関ルールなど)および結果の可視化から構成される(フェルドマン等, 2010)。

近年では、欧米諸外国の動向を踏まえて、専門家の間では、テキスト・マイニングという呼び名を改め、テキスト・アナリティクスと呼ぶことが提唱されている(言語理解とコミュニケーション研究会(NLC研究会)第10回テキストマイニング・シンポジウム2017.2.9-10)。しかしながら、世間的にはテキストマイニングの呼び名が定着しているため、本稿でも以降はすべてテキストマイニングで統一する。

先述のフェルドマン等は、テキストマイニングの目的は、情報源に潜在する興味深いパターンを試行錯誤しながら見つけ出し有用な情報を得ることであるとしている(フェルドマン等, 2010)。テキストマイニングでは、情報源として文書集合(corpus:コーパス)を利用する。文書集合は基本的に構造化されていない。テキストマイニングでは、構造化されていないテキストデータから、興味深いパターンを見つけて出す必要がある。この点が数値データを分析する場合と異なり、テキストデータの分析を難しくしている。

テキストデータとは、画像や動画、音声等ではない文字によって記録されたデータのことである。文字を厳密に言えば、数字も含まれる。なお、テキストデータのことを、テキストファイルと呼ぶこともある。テキストファイルは、バイナリファイルの対義語である。テキストファイル(テキストデータ)は、基本的に人間が目で見えて判読できるものである。一方のバイナリファイルの場合は、人間の目を見た場合に理解できない文字列であったり、モニター上に正しく表示されなかったりするものである。紛らわしいため、本稿では、数字を主体とするテキストデータのことを数値データと呼び、それ以外の非数値の文字が主体のデータのことを、テキストデータと呼び分ける。

従来の統計処理では、基本的に数値データを取り扱ってきた。また、近年の情報爆発のなかで生

3) その反面、テキストマイニングには一般的な定義がないという指摘もある(喜田, 2008)。その理由は、テキストマイニングは、関連領域、方法論、適用範囲が多様すぎるためである、としている。

成された大量の数値データを分析することを、データマイニングと呼ぶようになった。数値データに対する、データマイニング処理では、データは構造化された形式で蓄積されていることを前提としている。従って、データ分析の際の前処理として、データクレンジングと呼ばれるノイズデータを除去する作業が必須である。

テキストマイニングは、データマイニング分野の影響を強く受けている。テキストデータは数値データに比べると構造化されていないために、データマイニングとは異なった処理が必要になる。その処理が、非構造化データであるテキストデータを、構造化された中間表現に変換する処理である。また、テキストデータは自然言語で書かれている。自然言語とは、日常われわれがコミュニケーションをおこなうために用いる言葉である。そのために、次節で述べるように、自然言語処理と呼ばれる研究領域によって開発されてきた技術基盤を用いることが必要になってくる。

2-2 自然言語処理

われわれは日常読み書きや、会話に言語を用いている。それらは、日本語であったり英語であったりする。こうした言語のことを自然言語と呼ぶ。自然言語処理 (natural language processing : NLP) とは、自然言語を処理する技術や学術分野の総称である。自然言語処理は、人工知能研究と言語学研究的の重なり合った研究領域である。処理に当たっては、コンピュータに処理させる。基本的には文書として蓄積されたテキストデータの分析が中心である。近年では、音声データの処理についての研究も進んでいる。

自然言語処理が対象とするおもな処理内容は、文書分類 (document categorization), 機械翻訳 (machine translation), 文書要約 (text summarization), 質問応答 (question answering), 対話 (dialog) などである (平井他, 2017)。さらに、パソコンでの日本語入力ソフトや Web 上での検索エンジンなどの処理も対象としている。さらに、スマートフォンにおける音声入力によるバーチャルアシスタント (AI アシスタントとも言う) は、自然言語処理を用いて、質問に答える、推薦する、Web サービスを利用する、などの処理を行う。代表的なサービスとしては、Apple の siri, Google の Google アシスタント, Amazon の Alexa, Microsoft の cortana などがあり、それぞれ広く普及している。このように自然言語入力は、既にわれわれの日常生活の中で広く応用されている。また、今後も音声認識や音声合成など、さらに応用領域が広がることが予想されている。

自然言語処理は、機械翻訳を中心として発展してきたとされている。機械翻訳研究は、ロシア語やフランス語を英語に翻訳することを目的として、米国において始められた。1960 年代の米ソ冷戦時代には、ソ連の科学技術に対抗することを目的として、ロシア語から英語に翻訳するシステムの研究が米国で盛んに行われた。ロックフェラー財団のウォーレン・ウィーバーがノバート・ウォナーに当てた手紙の中で、ロシア語から英語への機械翻訳のアイデアが述べられている。手紙の中では、①翻訳を暗号の問題として捉える、②もともとは英語で書かれた内容が奇妙な文字 (筆者注キリル文字) によって暗号化されている、③暗号を復号化することによって機械翻訳ができる、というアイデアである。また、人工知能への応用も進められた。1960 年に開発された関数型言語 LISP を用いて、野球に関する質問応答システムが開発された (小町, 2016)。

日本における自然言語処理は、1980 年頃から活発に進められた。当初の目的は、日本の科学技術を英語に翻訳する日英翻訳システムの開発であった。また、当時商品化されたワードプロセッサ

向けに日本語入力システムの開発が進められた。日本語の場合、アルファベットだけを用いる英語と異なり、ひらがなや漢字等も用いるため、文字種が非常に多い。そのために、日本語入力技術の研究が進められた。ジャストシステムの日本語入力システム（ATOK）などが製品化された。さらに、日本語の自然言語処理では、形態素解析や、係り受け解析などの研究が進み、2000年代にChaSenをベースにしたMeCabが開発された。MeCabはCRF（条件付き確率場）によってコスト推定する汎用テキスト処理エンジンであり、今日でも広く利用されている（小町，2016）。

1990年代に入って、後述の統計的自然言語処理と呼ばれる手法が登場し、自然言語処理の分野に大きな転機となった。統計的自然言語処理とは、確率・統計モデルを用いた自然言語処理である。この背景には次の二つの要因がある。一つは、Webの発展にともなうビッグデータ環境の出現により、取り扱うテキストデータの量が膨大になったこと。もう一つは、確率・統計論や機械学習の手法が洗練されてきたことである。

最後に、自然言語処理の主要な処理内容について述べる。自然言語処理には、その目的（文書分類や文書検索等）を達成するために、下位の課題としていくつかの主要な言語解析タスクがある。言語解析タスクの代表的なものとして、坪井等（2017）は以下をあげている。

- 品詞タグ付け：単語に名詞・動詞などの文法的な役割処理（品詞）を付与する処理
- 単語分解：日本語など単語に分けられていないテキストを単語列に分割する処理
- 語義曖昧性解消：複数の語義を持つ単語の語義を特定する処理
- 固有表現抽出：人名・知名・日付などを抽出する処理
- 構文解析：文法に基づく文の木構造を構築する処理
- 述語項構造認識：述語を中心とした意味構造を抽出する処理

言語解析タスクの発展は新しい問題を生み出した。言語解析タスクは、「部分問題へと最適化されて」それぞれ洗練と発展してきた。それと同時に各タスクは相互に複雑な依存関係を持っている。その結果、自然言語処理を実践するさいに、全体に向けた最適化が困難になっている。複雑化した自然言語処理を人手で行うことが困難になっているからである（坪井，2017）。

人手で困難な処理を行うのが機械学習である。ビッグデータを用いたデータマイニングの際には、機械学習は必須である。テキストマイニングも同様である⁴⁾。自然言語処理において機械学習が用いられるようになったことは、自然の流れであるといえる。また、先述の、自然言語処理の転機となった統計的自然言語処理には、機械学習の利用が欠かせない。機械学習の概要を次節にまとめる。

2-3 機械学習

機械学習（Machine Learning）とは、データを分析する際に、コンピュータのアルゴリズムを用いて特長となるパターンを抽出することである。機械学習はオープンソースのPythonの普及に

4) 坪井等（2017）は自然言語処理では機械学習では不十分でさらに深層学習（ディープ・ラーニング）の活用が重要であると指摘している。本稿では、深層学習までは含めないものとする。

よい、広く用いられるようになった。さらに、Python には scikit-learn のような機械学習用ライブラリが存在している。このライブラリには、機械学習の主要なアルゴリズム（サポートベクターマシン、ランダムフォレスト、クラスタリング手法、ロジスティクス回帰等のさまざまなアルゴリズム）が含まれており、機械学習利用の普及に拍車が掛かっている。

機械学習は、取り組む問題の性質によって、大きく①教師あり学習、②教師なし学習、に分類される⁵⁾。教師あり学習を用いる場合は、識別（classification）や回帰（regression）を目的とする場合である。教師なし学習の場合は、クラスタリングやパターンマイニングを目的とする。また、機械学習の際には、従来の統計学に加えてベイズ統計の考え方も用いられる。

2-4 テキストマイニングの基本的な手法

テキストマイニングはコンピュータでおこなう。従って、テキストデータの言語表現をコンピュータが扱うことが出来るように、数値化する必要がある。そのために、文書、文、単語と言ったテキストデータの構成要素を数値表現に変換する処理を実行する。自然言語処理における、テキストデータの数値化処理を実行する際に基盤となっている主な概念は、以下の通りである。

○ トークン（token）とタイプ（type）

テキストデータの数値化処理の基本は、文書中に出現する単語を数え上げることである。その際に、一つ一つの単語の出現を示す用語がトークンである。それに対して、出現する単語の種類の数を示す用語がタイプである。それぞれ、単語トークン、単語タイプと呼ぶこともある。

○ N-gram, 連なり, 共起

テキストマイニングでは、隣接する複数の単語の連なりを単位に分析をおこなう。N 個の要素の連なりのことを N-gram と呼ぶ。1 つだけの要素（1-gram）の場合は、モノグラムあるいはユニグラムと呼ぶ⁶⁾。2 つの要素のつながりは 2-gram（バイグラム）、3-gram（トライグラム）は 3 つの要素の連なりである。このように、文章に含まれる文字や単語のつながりを対象として分析を行う。

分析の際には共起に着目する。共起とは文章の中で同時に生起する 2 つの要素のペアを集めたものである。一般に共起している場合はそれらの個々の間に何らかの意味的なつながりがあると考えられる。共起ペアの出現頻度が高ければ、文書中におけるそのペアの重要度が高いと言うことになる。共起ペアを線でつないだ図を書いて、視覚的に解釈する場合もある。

○ TF-IDF（Term Frequency- Inverted Document Frequency）

TF-IDF とは、文書中に出現する単語の頻度を、重みを付けて調整する手法である。テキストデータの分析の基本は、出現する単語の出現頻度をカウントすることである。ただし、出現する回数が多いからと言って、その単語が必ずしも文書の中で重要な単語であるわけではない。頻発する

5) なお、③中間的学習（半教師あり学習と強化学習）については、本論文では省略する。

6) ただし、モノグラムの出現回数については文字の単語レベルの出現頻度を数えることと同じであるため意味がない。

一般名詞（たとえば、「もの」とか「こと」のような単語）は、出現数は多くても、本当に重要なキーワードではない。特定の文書にのみ頻出する単語に注目するほうが、テキストデータ分析では有効である。こうした点を修正するために、生の頻度を調整して重み付けする必要がある。

TF (Term Frequency) とは、単語の出現回数である。また、IDF (Inverted Document Frequency) とは、その語が出現する文書数の逆数の log をとった数値である。一部の文書に出現する単語に対して重みを与える。TF と IDF を掛け合わせた TF-IDF を用いることによって、その単語の重要性を図る指標とする。TF-IDF の値を用いて文書中のキーワードを抽出したり、それ以外の分析処理の前処理として、重要な単語のみを選択（フィルタリング）処理をしたりする。

○ 文書のベクトル表現と bag of words (BOW)

文書を分類するためには、文書特有の特徴を利用する。こうした特徴のことを素性 (feature) と呼ぶ。また、素性の持つ値のことを素性値 (feature value) と呼ぶ。ベクトル空間モデル (vector space model) では、文書の素性値として単語の頻度を数え上げ、文書をベクトルとして表現する。文書同士の内容の類似度を測定するために、コサイン類似度などの手法でベクトル空間上で比較する。

文書のベクトル表現としてよく知られている手法が、bag of words (BOW) である (図 1)。BOW では、文書内に出現するすべての素性 (単語) を用いてベクトル化する。その際に、二値、頻度、TF-IDF 値等を用いる。BOW は、単語をバラバラに切断して、一つの袋に入れた状態であるため、語順などの情報は喪われてしまう。また、単語すべてを用いるために、ベクトルの次元が巨大になってしまう。この点を解消し、ベクトルの次元を 100-200 程度に落とすためのアルゴリズムが、次に述べる Word2VEC である。



図 1 BOW 図

出典：岩田, 2015 p. 18 より

○ Word2Vec, Doc2Vec

統計的自然言語処理は、さまざまな機械学習アルゴリズムによって研究が進められてきた。そうしたアルゴリズムの一つがニューラルネットワークである。ニューラルネットワークは、1940年代から研究されている、古いアルゴリズムの一種である。一時期研究が停滞した時期もあったが、近年また見直され、今日隆盛しているディープ・ラーニングの基盤となっている。自然言語処理の中でも、ニューラルネットワークにもとづいたアプローチが盛んに研究されている。ニューラルネットワークを用いた Word2Vec というアルゴリズムが使われるようになってきた。

Word2Vec は、数十万次元になる BOW によるベクトルを、100 もしくは 200 次元に縮小するも

のである。Word2Vec がうまく機能する理由は、ニューラルネットワークを用いて意味を伴う次元圧縮をおこなっているからだと言われている。最近では、Word2Vec の技術をベースに、単語だけではなく文書を対象とし、「文書」に意味を持たせてベクトルとして捉えて利用する、Doc2Vec などが用いられている。また、一定の規則を持って連続的に並んでいるものであれば応用できるため、Web ページの遷移をもとにしたレコメンデーションや、アイテムの推奨 (Item2Vec)、画像処理 (Image2Vec) などとも試みられている。

2-5 テキストマイニングの応用

テキストマイニングの手法は、さまざまな応用が期待できる。まず、営業員が記録する営業日報を分析することが考えられる。営業日報は、SFA ツールが普及して業務報告書のデジタル化・スコア化が進んでいる今日でも、企業の業務管理の現場で依然として広く用いられている。その反面、蓄積された文書データはほとんど分析されず、あまり有効活用されていない。その理由は、文書データが構造化されていないためデータ処理が難しいからである。埋もれたままの営業日報という膨大な文書データを分析することが出来れば、貴重な知見を得ることが出来るであろう。また、顧客からのクレームやコメント、SNS への書き込みなどの分析も有効であると期待できる。ビジネス活動を通じて蓄積されているさまざまなテキストデータについての分析が、今後ますます高度化されていくであろう。

3 統計的自然言語処理

統計的自然言語処理の中で、比較的新しい分析手法であるのが潜在意味解析と呼ばれるアプローチである。本節では、潜在意味解析について述べる。

3-1 潜在意味解析

テキストデータを分析するためには、本来であれば文書に含まれる単語の意味を解釈する必要がある。コンピューターに単語の意味を分析させることは容易ではない。そのための辞書の整備にも多大な労力が要求される。また、大量のテキストデータを分析するためには、新しい分析アプローチも求められてきた。そうした中で開発された新しい分析手法が、統計的潜在意味解析と呼ばれるアプローチである。統計的潜在意味解析 (Statistical Latent Semantic Analysis) とは、テキストデータが持つ意味を数学的に表現して分析することである。統計的潜在意味解析は、表 1 のように歴史的に発展してきた。

表 1 潜在意味解析の変遷

年度	名称	略称	英語名称	特徴・手法
1988	潜在意味解析	LSI/LSA	Latent Semantic Indexing/Analysis	行列分解
1998	確率的潜在意味解析	PLSI/PLSA	Probabilistic LSI/LSA	確率モデル
2003	潜在ディリクレ配分法	LDA	Latent Dirichlet Allocation	階層ベイズモデル

奥村／佐藤 (2015) p.10 より抜粋し著者作成

潜在意味解析 (Latent Semantic Indexing/Analysis) では、文書を単語の出現頻度でベクトル化する。ベクトル化した単語を並べて行列を作り、特異値分解 (singular value decomposition) と呼ばれる行列分解によって、ランクや次元を削減し圧縮する技術である。潜在の意味とは、文書の背景に潜在するトピックのことである。このトピックを抽出し、文書の分析をおこなう。

潜在意味解析の理解のために、例えば、「車を運転する」という文書と「自動車を運転する」という二つの文書があったとする。「車」「自動車」という単語のみに注目すると、この二つの文書が似通った内容であるとは判断出来ない。なぜならば、これら二つの単語が文書中で表層上は共起していないからである。ところが、潜在意味解析をおこなうことによって、「車」「自動車」と共起する「運転する」という単語を通じて、「車」「自動車」の間の潜在的共起性を抽出することが可能となるのである。

潜在の意味解析の背景には、1950年代に発表された、ハリスの分布仮説 (distributional hypothesis) がある (Harris, 1954)。分布仮説とは、単語の意味の解釈は、その単語が使われた文脈 (周辺の単語) によって決まる、という考え方である。つまり、同じ文脈に出現する単語は同じ意味を持つ、という考え方である。この考え方をを用いることによって、文書の意味合いを統計的に分析し、文書の類似度を判断することが可能になる。

3-2 確率的潜在意味解析と潜在ディリクリ配分法 (LDA)

潜在意味解析の考え方を発展させたのが、確率的潜在意味解析 (Probabilistic LSI/LSA) である。このアプローチは、文書の背景に複数のトピックが存在していると想定する。文書に含まれる単語は、トピック固有の単語出現確率 (単語分布) によって生み出されると考える。そして、収集したテキストデータから、元になっている確率分布を最尤推定する手法である。

潜在ディリクリ配分法 (Latent Dirichlet Allocation : LDA) は、確率的潜在意味解析 (pLSI/pLSA) を階層ベイズモデルとして発展させたものである。確率的潜在意味解析 (pLSI/pLSA) では文書のトピック混合比率 (トピック分布) は固定されているが、潜在ディリクリ配分法 (LDA) ではトピック分布も確率的に生成されると考える。その上で、事前分布と事後分布にディリクリ分布を想定して、ベイズ推定によって単語分布とトピック分布を推定する手法である。文書の背景に隠されているトピックを推定し考察することにより、新しい知識を獲得することが可能となる。

3-3 トピックモデルとグラフィカル表現

確率分布による生成過程を想定した確率モデルのことを、総称してトピックモデルと呼ぶ。データの背景に隠れたトピックという潜在変数を抽出することにより、潜在変数が持つ情報から知識を得ることが期待できる。トピックモデルは本論文で述べた文書データの分析以外に、画像処理、購買記録データの分析、推奨システムなど、多くの分野で用いられている。トピックモデルのように、データの生成過程を確率変数を用いて表現したものを確率生成モデル (probabilistic generative model) もしくは生成モデルと呼ぶ。生成モデルを図示する場合には、グラフィカルモデル表現 (graphical model representation) が用いられる。グラフィカルモデルによって表現すると、モデル内の変数の関係が視覚的に把握でき、生成過程がわかりやすくなる。トピックモデルのバリエーションとしては、ユニグラムモデル、混合ユニグラム、トピックモデルがあり、後に行くほど

複雑なモデルとなっている。図2は、三つのモデルのグラフィカルモデル表現である。

- ユニグラムモデルでは、すべての単語は一つのカテゴリカル分布に従って生成されると仮定している。
- 混合ユニグラムモデルでは、トピック毎に異なる単語分布を持っている。トピック分布については、文書全体で一つ持っている。
- トピックモデルでは、一つの文書が複数のトピックを持つと仮定している。文書毎にトピック分布を持っている点が、混合ユニグラムモデルと異なっている。

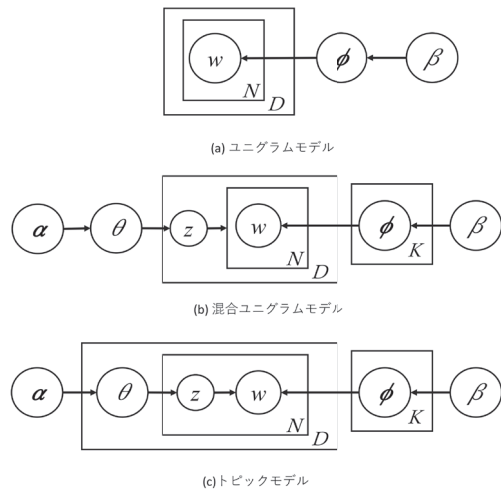


図2 グラフィカルモデル表現

出典：岩田，2015 p.58

凡例：D：文書数，N：文書に含まれる単語数，K：トピック数，W：単語，z：トピック， θ ：トピック分布， ϕ ：単語分布， α, β ：ハイパーパラメーター

4 営業研究について

4-1 営業研究の必要性と現状

企業経営にとって営業が重要であることと、営業を取り巻く状況が変化していることから、営業についての研究が必要を増している。しかしながら、営業活動に対する研究は我が国においてはあまり活発におこなわれてきたとは言いがたい⁷⁾。日本国内では営業関連の実用書は多く刊行されているが、個人の体験に基づく成功譚を展開しただけのレベルのものが多く。学術書では石井等(1995)、田村(1999)、高嶋(2002)によるものがある程度である。

こうした背景として考えられることは、営業活動に関する研究が学問として体系的に整理されおらず、学問領域として確立されていない状況があげられる。学問上の分類としては、営業管理はマーケティングの範疇に入っている。いわゆるマーケティングの4Pの一つであるプロモーションのうちの一つの要素にとどまっている。そのため、研究課題が矮小化されたものに限定されていると指摘されている(石井, 2012)。

海外では営業研究は盛んに行われている⁸⁾。Journal of Marketing, Journal of Marketing Research, Marketing Science等の有力雑誌の他、営業研究に特化した、Journal of Personal Selling & Sales Management (JPSSM) という学術誌が季刊で発行されている。JPSSMの研究テーマの動向としては、売手と買手の関係、営業員の業績と満足度の関係、組織内での営業管理といった内容が多かった。近年では、SFA、ITとデジタル化、営業員のストレスといった内容が注目されている。営業員の資質と個人業績に関する研究も多い。

7) 営業研究の現状については、以前から言及している(北中2018, 2019)。

8) Richards等(2010)は、営業研究における米国人以外の研究者が近年急増していると指摘しているが、それでもアジア系の研究者の割合は0.6%にとどまっているのが現状である。

営業研究関連の国際的な学会としては、2007年に結成されたThe Global Sales Science Institute (GSSI)⁹⁾が活発な活動を展開している。この学会はBtoBにおける営業管理の諸側面を対象としており、世界中の学者と実務家のメンバーによる、情報交換と交流をおこなっている。毎年6月にGSSI Annual Conferenceを世界各地で開催している。それ以外では、AMA (American Marketing Association)の下部組織であるAMA Sales Special Interest Group (SalesSIG)が米国で年次大会を開催している。

営業に関する教育についても同様、海外の方が活発である。ここでいう営業教育とは、実務家向けのセミナーの類いではない、アカデミックな内容の教育を指している。米国においては大学間の営業教育のコンソーシアム (USCA: University Sales Center Alliance) が結成されており、研究活動と教育、就職支援等の活動を展開している。ヨーロッパにおいても、フィンランド、ドイツの大学において営業教育が提供されている。

日本国内では、プルデンシャル生命保険が2015年度より「営業」をテーマとした寄付講座・講義を開講している。現在では、東北大学(経済学部)、青山学院大学(全学部共通)、慶應義塾大学(商学部)、中央大学(商学部)、立命館大学(経済学部)の5大学で開講している。2019年度からは、早稲田大学(商学部)、明治大学(全学部共通)、同志社大学(商学部)を加えた8大学に拡大している。わが国における営業教育の広がりの兆しがうかがえる。

4-2 営業研究の先行研究レビュー

JPSSMには、時々文献研究論文が投稿され掲載されている。本研究では、過去10年間の間にJPSSMに掲載されたレビュー論文について、主なものの概要を以下にまとめておく。

レビュー論文の中でも、その内容に違いがある。まず、過去の一定期間の営業研究全論文を対象に研究内容の傾向を調べたものがある。Asare等(2012)は、1980年から2008年までの29年間において、主要15ジャーナルに掲載された1,346本の営業研究論文に対して、研究手法のトレンドに関する分析を行っている。Schrock等(2016)は、JPSSMに掲載された721本の論文の中にある35,975件の引用に対して、計量書誌学的分析(Bibliometric Analysis)をおこなっている。その際に、1980年代、1990年代、2000年代の3世代に分けて、多次元尺度構成法(multidimensional scaling)によって、主要な研究テーマの抽出し、それぞれの世代毎の傾向分析を試みている。

次に、営業研究のうち特定のテーマに絞って傾向を分析したものがある。Gustafson等(2018)は、組織風土と文化に関する既存の組織研究から、営業研究に関連する概念的基盤を明らかにしている。Khusainova等(2018)は、動機についての先行研究から、動機の定義、動機の基盤となっている主要な理論、動機の測定方法、等について整理した。その上で、セールスパウソンの動機付けに影響を及ぼす要因を考察している。また、セールスパウソンの動機に関する先行研究の一覧を提供している。Goad等(2014)は、1982年から2013年の間の、売手志向と顧客志向と営業成績に対する影響度合いについて、メタアナリシスによって先行研究を分析した。Richards等(2010)は、2003年から2007年の間に19のマーケティング関連ジャーナルに掲載された388本の営業研究について、過去の同様の調査(1993-1997, 16ジャーナル, 266本)との比較をおこなった。

9) <http://gssi.world/>

Panagopoulos 等 (2011) は、営業活動の国際化によって国際営業研究が増加していることを受けて、1980 年から 2009 年の間に発表された国際営業研究を分析した。

また、著者の共著関係をネットワーク分析したのも出てきている。Yang 等 (2010) は、1980 年から 2009 年の間、JPSSM に掲載された 628 本の論文とその著者 761 名に対して、共著関係をもとにしたネットワーク分析をおこなっている。Schrock 等 (2018) は、1980-2017 年の 132 本の国際営業研究についてネットワーク分析をおこなった。関連する用語の分析を通じて、研究内容の相互関係を洗い出している。

4-3 分析対象ジャーナルについて

テキストマイニングを適用するに際して、営業研究に特化した学術専門誌の掲載論文を対象とした。その目的は、営業研究分野における研究動向を抽出することである。本研究では、2016-2018 の 3 年間の間に、Journal of Personal Selling & Sales Management (JPSSM) に掲載された 63 件の論文 (レビューや Editorial 的なものは除いた一般的な論文) を対象として、テキストマイニングによる分析をおこなった。分析の対象としては、論文に掲載されている Abstract を対象とした。

4-4 データ分析の手順

潜在的意味解析である潜在ディリクリ配分法 (LDA) による、アブストラクトの分析にあたっては、以下の処理を施した。

① トークン化 (tokenization) / 分かち書き

トークン化とは、文書から文の要素 (単語) を切り出すことである。日本語の場合は分かち書きと呼ばれる。テキスト処理の基本である。英語の場合は、文章中で単語単位で既に別れているため処理は容易である。日本語の場合は、おもに助詞を手がかりにする。切り分けた単語を、リスト形式でファイルに保存する。

② 不要語 (stop word) の除去

切り分けられた単語から、テキスト解析をおこなう上で不要なものを除去する。こうした不要語は、出現頻度が多いわりに文章の意味にはほとんど影響しない。日本語の場合、文末の「です」「である」のような単語が該当する。英語の場合には、冠詞、前置詞、接続詞などが該当する。

③ 見出し語化 (レンマ化) / lemmatization

切り分けた単語を解析して、辞書の見出し (lemma) の形に戻す処理である。品詞を特定し、活用形や原型をもとめる処理をおこなう。日本語の場合は、形態素解析と呼ぶ。形態素解析ソフトウェアの代表的なものが、オープンソースの MeCab である。英語用の Python から使える自然言語処理ライブラリとしては、本研究で使用した NLTK (Natural Language Tool Kit)¹⁰⁾がある。

10) <http://www.nltk.org/>

④ 語幹化 (stemming)

英語の場合、語幹が共通で語尾が変化する単語がある。こうした単語については、文書中で同一の単語として扱う方が解釈には有効な場合もある。例えば、swim という単語の場合、「swims」「swimming」「swimmer」などの複数の語形があるが、すべて意味するところは swim で事足りる。こうしたバリエーションを取り除く処理である。なお、本研究では lemma 化で充分であったため、適用しなかった。

データ分析をおこなった環境は、下記の通りである。

- OS : Window10/Ubuntu 18.04.3 LTS
- プログラミング言語 : python 3.7.4
- 言語処理ライブラリ : NLTK 3.4.5
- LDA ライブラリ : gensim 3.8.0¹¹⁾

4-5 分析結果と考察

データ分析の結果が以下の表である。分析には gensim の LDA モデルを用いた。トピック数を 5 と指定した分析結果である。なお、それぞれのトピックに対して、20 のキーワードを出力させた。

表 2 分析結果

抽出されたトピック	割り当てられた単語			
0	0.025**"manager" 0.013**"role" 0.011**"relationship" 0.010**"mia" 0.008**"reward"	0.015**"business" 0.012**"performer" 0.010**"perceived" 0.010**"performance" 0.008**"key"	0.014**"team" 0.011**"level" 0.010**"share" 0.009**"management" 0.007**"intelligence"	0.014**"control" 0.011**"understanding" 0.010**"salesperson" 0.008**"future" 0.007**"collaboration")
1	0.049**"customer" 0.012**"buyer" 0.010**"social" 0.008**"benefit" 0.008**"tie"	0.017**"relationship" 0.011**"performance" 0.009**"role" 0.008**"salesperson" 0.008**"team"	0.014**"behavior" 0.010**"service" 0.008**"interaction" 0.008**"gratitude" 0.007**"model"	0.013**"marketing" 0.010**"perception" 0.008**"leadership" 0.008**"integration" 0.007**"selling"
2	0.020**"customer" 0.014**"service" 0.010**"business" 0.008**"provide" 0.008**"outcome"	0.019**"selling" 0.011**"relationship" 0.009**"motivation" 0.008**"context" 0.007**"organizational"	0.015**"climate" 0.010**"function" 0.009**"strategic" 0.008**"finding" 0.007**"specific"	0.014**"management" 0.010**"marketing" 0.009**"role" 0.008**"key" 0.007**"theory")
3	0.020**"data" 0.014**"effort" 0.011**"issm" 0.009**"network" 0.007**"cultural"	0.020**"performance" 0.014**"negotiation" 0.009**"employee" 0.009**"panel" 0.007**"technology"	0.018**"organization" 0.013**"structure" 0.009**"role" 0.009**"resistance" 0.007**"key"	0.018**"outcome" 0.013**"online" 0.009**"relational" 0.007**"survey" 0.007**"seller"
4	0.019**"emerging" 0.013**"attachment" 0.011**"listening" 0.009**"data" 0.009**"strategy"	0.019**"customer" 0.013**"finding" 0.011**"job" 0.009**"effect" 0.008**"future"	0.017**"market" 0.013**"selling" 0.010**"developed" 0.009**"model" 0.008**"analysis"	0.015**"performance" 0.011**"influence" 0.010**"relationship" 0.009**"brand" 0.008**"capability"

11) <https://radimrehurek.com/gensim/>

簡単に結果を見ると、トピックを以下のように解釈できる。ここで、トピックとは、対象となった論文に共通する話題（研究課題）であるとみなせる。また、一本の論文が複数の話題を持っている場合もありえる。それぞれのトピックの内容を説明し、一言に集約するキーワードを段落の最後の【 】内に表記した。

- トピック0は、販売活動に関してチームやマネジャーのあり方に関するものである。営業研究の中心的なテーマの一つであるといえる。【営業管理】
- トピック1は、顧客、マーケティング部門、営業員との関係についてである。相互作用やリーダーシップという用語が出現しており、具体的な研究内容が伺える。【関係性】
- トピック2は、組織（機能、サービス、風土等）と顧客に対する戦略（考え方）に関するテーマであると理解できる。【戦略志向】
- トピック3は、少し解釈が難しいが、さまざまな要素と業績の関係についての分析手法に関するテーマであると考えられる。【業績分析】
- トピック4は、新興市場における営業活動についてである。近年、ブラジルやメキシコといった南米における営業活動の研究が増えており、それが反映されている。【新興市場】

今回の分析では、LDAによるトピックモデルによる分析の結果、「営業管理」、「関係性」、「戦略志向」、「業績分析」、「新興市場」という五つのトピックを抽出することが出来た。これらの五つのトピックは、筆者が63本の論文を購読して獲得した印象とおおよそ共通していた。従って、テキストマイニングによって、有意義な研究動向を抽出することは、可能であると考えられる。

次に、今回の対象論文は63本と少ないので、テキストマイニングによって分析をするよりも、普通のやり方（人力による手作業）の方が効率がいいかも知れない。しかし、これが数百本、数千本の論文となった場合には、人力による分析では大変な労力と手間が要求される。上述のように、研究期間や対象を拡大するような場合には、人力による分析では追いつかない可能性もある。そうした大量の文献を対象とした分析をおこなう場合には、本研究のような機械学習的アプローチのテキストマイニングは、解決策の一つになるだろう。また、コンピューターを使って機械的に分析した方が、人力による解釈よりも個人的な好悪や馴染みの有無等から生じるバイアスが少なく、客観的な分析ができる可能性があるすなわち、潜在的意味解析による文献分析は、対象領域の動向を価値フリーでバイアスが少なく抽出が可能となる点で有用性が高い。その結果、研究動向に関して、従来の手法より正確にトレンドを抽出できることが期待できる。

以下の点が、本研究の限界であり今後の課題である。まず、テキストマイニングの分析対象としたのが論文のabstractの文章である点の検討が必要であるだろう。Abstractは明示的ではないが分量に制約がある。そのため、表現方法に紋切り型の表現が用いられ書き方に片寄りがある可能性がある。そのため、トピックの切り出しがあまりうまく出なかった可能性もある。また、論文で特集号が組まれた場合には、同じテーマの論文が集まる。そのため、内容に片寄りが生じ、有効なトピックが抽出出来ない可能性もある。

今回の分析では、あくまでも試行という位置づけでおこなった。そのため、対象期間は3年間に絞り込んだため、対象とした論文数は63本と少ないものであった。今後、本格的にテキストマイ

ニングによる営業研究の文献研究をおこなうためには、対象を広げた分析が必要であろう。例えば、Journal of Personal Selling & Sales Management (JPSSM) 誌の創刊から今日までの全論文を対象とことや、JPSSM 誌以外の論文誌に掲載された、営業研究論文を分析対象とすることなども検討すべきであろう。

5 結 論

本研究は、文献研究法に機械学習アプローチの一種であるテキストマイニングの適用を試みたものである。テキストマイニングを適用するに際して、営業研究に特化した学術専門誌の掲載論文を対象として取り上げた。テキストマイニングの目的は、営業研究分野における研究動向を抽出することである。その際、確率的潜在意味解析（潜在ディリクリ配分法：LDA）という手法を用いた。LDAによるトピックモデルによる分析の結果、「営業管理」、「関係性」、「戦略志向」、「業績分析」、「新興市場」という五つのトピックを抽出した。

潜在的意味解析による研究テーマの動向分析は、①大量のデータを対象とした分析が可能である点、②人手を介さずに分析をおこなうのでバイアスのない研究テーマ動向の抽出が可能である点、から有意義である。文献研究にテキストマイニングを用いることの有用性は高く、今後盛んになる可能性がある。

謝 辞

本研究は、JSPS 科研費 17 K04003 の助成を受けたものである。

参考文献

1. 石井淳蔵, 『営業をマネジメントする』, 岩波書店, 2012 (石井淳蔵, 『営業が変わる』, 岩波書店, 2004 の改題・復刊)
2. 石井淳蔵・嶋口充輝, 『営業の本質』, 有斐閣, 1995
3. 岩田具治, 『トピックモデル』, 講談社, 2015
4. 大木秀一, 『文献レビューのきほん』, 医歯薬出版, 2013
5. 奥村学 (監修)・佐藤一誠, 『トピックモデルによる統計的潜在意味解析』, コロナ社, 2015
6. 奥村学 (監修)・高村大也, 『言語処理のための機械学習入門』, コロナ社, 2010
7. 喜田昌樹, 『テキストマイニング入門——経営研究での活用法』, 白桃書房, 2008
8. 北中英明, 『法人営業・営業員の特性別要因による分類』, 経営経理研究 (拓殖大学紀要), Vol. 112, 2018/3, pp. 257-277
9. 北中英明, 『セールスパーソンの資質分析——ビッグファイブ性格特性によるアプローチ』, 経営経理研究 (拓殖大学紀要), Vol. 115, 2019/3, pp. 23-40
10. 北中英明, 『社会科学における研究方法について』, 経営経理研究 (拓殖大学紀要), Vol. 61, 1998/12, pp. 181-203
11. G. キング, R. O. コヘイン, S. ヴァーバ, 真淵勝, 『社会科学のリサーチ・デザイン』, 勁草書房, 2004
12. J. W. クレスウェル, V. L. プラノ・クラーク, 『人間科学のための混合研究法』, 北大路書房, 2010
13. 小町守 (監修)・奥野陽・グラム=ニュービッグ・萩原正人, 『自然言語処理の基本と技術』, 翔泳社, 2016
14. 高嶋克義, 『営業プロセス・イノベーション』, 有斐閣, 2002
15. 田村正紀, 『機動営業力』, 日本経済新聞社, 1999
16. 坪井祐太・海野裕也・鈴木潤, 『深層学習による自然言語処理』, 講談社, 2017
17. 樋口耕一, 『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, 2014
18. ヘンリー・ブレイディ, デヴィッド・コリアー, 『社会科学の方法論争』, 勁草書房, 2008
 19. ウヴェ・フリック, 『質的研究入門』, 春秋社, 2011
 20. ローネン＝フェルドマン／ジェイムズ＝サンガー, 『テキストマイニングハンドブック』, 東京電機大学出版局, 2010
 21. 松村真宏・三浦麻子, 『人文・社会科学のためのテキストマイニング [改訂新版]』, 誠信書房, 2014
 22. 山内長承, 『Python によるテキストマイニング入門』, オーム社, 2017
 23. Asare, A. K., J. Yang and T. G. B. Alejandro (2012). “The State of Research Methods in Personal Selling and Sales Management Literature.” *Journal of Personal Selling & Sales Management* 32(4): 473-490.
 24. Goad, E. A. and F. Jaramillo (2014). “The good, the bad and the effective: a meta-analytic examination of selling orientation and customer orientation on sales performance.” *Journal of Personal Selling & Sales Management* 34(4): 285-301.
 25. Gustafson, B. M., N. Pomirleanu and B. John-Mariadoss (2018). “A review of climate and culture research in selling and sales management.” *Journal of Personal Selling & Sales Management* 38(1): 144-167.
 26. Khusainova, R., A. de Jong, N. Lee, G. W. Marshall and J. M. Rudd (2018). “(Re) defining salesperson motivation: current status, main challenges, and research directions.” *Journal of Personal Selling & Sales Management* 38(1): 2-29.
 27. Panagopoulos, N. G., N. Lee, E. B. Pullins, G. J. Avlonitis, P. Brassier, P. Guenzi, A. Humenberger, P. Kwiatek, T. W. Loe, E. Oksanen-Ylikoski, R. M. Peterson, B. Rogers and D. C. Weibaker (2011). “Internationalizing Sales Research: Current Status, Opportunities, And Challenges.” *Journal of Personal Selling & Sales Management* 31(3): 219-242.
 28. Richards, K. A., W. C. Moncrief and G. W. Marshall (2010). “Tracking And Updating Academic Research In Selling And Sales Management: A Decade Later.” *Journal of Personal Selling & Sales Management* 30(3): 253-271.
 29. Schrock, W. A., Y. Zhao, D. E. Hughes and K. A. Richards (2016). “JPSSM since the beginning: intellectual cornerstones, knowledge structure, and thematic developments.” *Journal of Personal Selling & Sales Management* 36(4): 321-343.
 30. Schrock, W. A., Y. Zhao, K. A. Richards, D. E. Hughes and M. S. Amin (2018). “On the nature of international sales and sales management research: a social network-analytic perspective.” *Journal of Personal Selling & Sales Management* 38(1): 56-77.
 31. Zellig S. Harris, “Distributional Structure,” *WORD*, 1954, 10: 2-3, pp. 146-162.
 32. Yang, Z., F. Jaramillo and L. B. Chonko (2010). “Productivity And Coauthorship In Jpssm: A Social Network Analysis.” *Journal of Personal Selling & Sales Management* 30(1): 47-71.

(原稿受付 2019年11月5日)