The Society for Economic Studies
The University of Kitakyushu
Working Paper Series No. 2011-12
(accepted in March 30, 2012)

日本におけるテキストマイニングの応用

齋藤朗宏 (北九州市立大学経済学部)

序論

言葉の分析においては，近年，テキストマイニングと呼ばれる研究分野が発展している．テキストマイニングとは，膨大なテキスト（文書）情報の中から有用な情報を掘り出す（マイニング）ことで，定型化されていないテキストデータを，一定のルールに従って定型化して整理し，データマイニングの手法を用いながら，相関関係などの定量分析を行う手法である．

文章の分析そのものには長い歴史がある．金 (2009b) によれば，19 世紀末には既に単語の長さの分布を用いた分析が行われている．自然言語テキストからの情報抽出についても，有村 (2003) によれば 1980 年代後半から研究されている．しかし，データマイニングの一手法としてのテキストマイニングという名が与えられ，特に実用化が進んできたのは，インターネットや PC の普及に伴い電子化テキストが急激に増加し始めた 1990 年代後半になってから (那須川, 2009) である．

ただ，その初期の研究は，主に理論研究と実用化のためのソフトウェアの発表が中心であり，応用研究は多くない．少数ではあるが見られた応用研究にしても，後述の那須川 (2001) を代表に，自分で必要なソフトウェアを開発するという方法を取られることが多かった．データマイニングの諸分野の中でも，応用が遅れているのは，テキストマイニングはある特定の言語への対応が求められるため，ある言語のために開発されたソフトウェアをそのまま他の言語に対して用いることができないという事情があったものと考えられる．最近では，樋口 (2004) の KH coder を嚆矢とし，松村・三浦 (2009) の TinyTextMiner，金 (2009a) の MLTP に見られるように，日本語を分析することのできるフリーのソフトウェアも豊富である．形態素解析ソフトウェア MeCab を R に組み込んだパッケージ RMeCab(石田, 2008) により，フリーの統計ソフトウェア R 上でもテキストマイニングは実行可能となっている．こういった背景から，最近では統計，テキストデータ分析の専門家ではない研究者による応用事例が数多く見られ，より一層の発展が期待されている．ただ一方で，那須川 (2009) が指摘するように，テキストマイニングという言葉やツールの普及と比べ，大きなインパクトにつながっている活用成功事例は少ない．Hearst (1999) の指摘する，貴重で新奇な情報を得てこそ真のテキストマイニングという立場からは，真のテキストマイニングに到達できていないとも言える．そこで，本論文では，日本におけるテキストマイニングの応用の現状を確認し，今後の発展の可能性について考察する．

テキストマイニングの技術

テキストマイニングの応用について考えるにあたり，テキストマイニングの基本的な考え方について解説する．テキストマイニングの入門書は数多く出ており，解説論文も少な

くない．中でも，松村 (2008) が全体の流れを理解するにはわかりやすいので，解説は同論文を基本として行う．

形態素解析

日本語のように単語間の区切りが明示されていない言語は，分析に先立って文章を分かち書きし，形態素に分割する．形態素とは，「言語学で，意味を持った最小の音型．ヤマ（山）のように形態素一つで単語が構成される場合もあれば，ヤマカゼ（山風）のように複数の形態素が単語を構成する場合もある（大辞泉）」とされる．文章から形態素を探し出し，その形態素単位に分割することを形態素解析と呼ぶ．日本語の形態素解析には, は京都大学黒橋研究室の JUMAN[*1]をはじめ，奈良先端科学技術大学院大学松本研究室の茶筌 (松本, 2000) や，google の工藤氏による MeCab[*2]など, フリーのソフトウェアがある．前述の KH coder，TinyTextMiner などでは，こういった形態素解析のツールを組み込んでいるので，ツールの存在を意識しなくても分析を進めることができる．

構文解析

形態素に分割された文章は集計の際には有効だが，文章の意味にまで踏み込んで分析を行う際には不十分である．こういった場合には，係り受け関係など構文について検討する必要がある．構文解析に用いられるソフトウェアとしては，京都大学黒橋研究室の KNP(黒橋, 2000) や，google の工藤氏による cabocha(工藤・松本, 2001) が挙げられる．これらは，JUMAN や MeCab 同様フリーソフトウェアである．

頻度集計

分析の第一段階は，単語の頻度の集計である．集計方法は，大きく分けて二通りある．一つは，文章の中で単語が出現した個数を集計する方法である．この方法では，一つの文章である単語が複数回出現した場合，それぞれを出現回数としてカウントする．もう一つは，文章の中で単語が出現したか否かを集計する方法である．この方法では，一つの文章の中である単語が何回出現したとしても，一回としてカウントする．単語の集計により，分析対象となる文章の特徴を大まかに把握することが出来る．

共起

意味の分析を考える場合，文章や段落内での共起関係の分析も有用である．これは，単語同士の分割表を作成する形で集計を行う．前述の頻度集計が一次元の集計であるのに対して，分割表の作成は，二次元の頻度集計と考えることもできる..

統計解析

テキストデータの場合，単語を用いて集計を行なっても，出現単語数が多くなりがちであり，そのため，単純に結果を見るだけでは，有効な知見を得るのは難しい．そこで行われるのが統計的手法を用いた分析である．たとえば，書いた人の性別や年齢といったテキストの属性と出現単語を用いたコレスポンデンス分析，あるいは個々の文章と出現単語を用いた数量化三類, それらの結果を用いたクラスター分析が考えられる．それ以外にも，単語間の共起性を見る多次元尺度法やネットワーク分析，また，SVM のような機械学習による，テキスト分類などもよく行われている．

テキストマイニングの国内における応用事例

経営学

最も多くの応用事例が見られたのは，経営学，あるいは経営の実場面においてであった．中でも，企業のカスタマーセンター，コールセンターにおける顧客とのやり取りにはごく初期から数多くの分析事例がある．中でも，那須川他 (1999)，那須川 (2001) は，テキスト

[*1] http://nlp.ist.i.kyoto-u.ac.jp/
[*2] http://code.google.com/p/mecab/

マイニングの応用研究の魁とも呼ぶことができ，数多く引用もされている．同論文では，問合わせからの概念抽出，係り受けパターンの分析，話題抽出や，時系列での話題の変化の分析などを，独自開発したソフトウェアを用いて行った例を示している．また，それらの成果は，FAQ として Web で公開されている．これ以外にも，長谷川 (2011) では長時間対応になったコールログについて，新人とベテランとの頻出係り受けの違いから，問題点を探り研修や FAQ を作成した．岡本他 (2000) では，問い合わせ電子メールの内容を分析し，返答例をオペレータに提示するシステムについて紹介している．また，上田他 (2004) や櫻井・酢山 (2005) では，問い合わせメールに対する分類手法についての提案，実例の提示が行われている．

カスタマーサポートとは異なるが，比較的近い応用先としては，営業日報の分析も挙げられる．櫻井・酢山 (2002) では，営業日報から，キーとなる概念を用いてテキストを分類する手法の提案をおこなっており，市村他 (2003) でも，営業日報を元に，成功事例，機会損失事例の分析を行い，要因と結果から，因果関係を持つ構造の抽出をしている．

一方近年では，マーケティング分野への応用が極めて多く見られる．豊田・森永 (2003) では，ブランドイメージの調査結果を用いて，ブランドとキーワードとのコレスポンデンス分析を行ないブランドのイメージを可視化している．黒岩 (2005) でもブランドイメージの調査結果からキーワードの集計を行なっている．石川・星野 (2004) では，観光地での落書き帳の書き込みからニーズを把握するために，キーワードの集計を行い，キーワード間の共起関係をスプリング埋込みという手法で確認している．小木 (2005) では，映画ジャンルごとの感想の頻出語句を確認し，ジャンルと語句のコレスポンデンス分析や係り受けの分析を行なっている．これ以外にも枚挙に暇がないため，主要なものについて箇条書きで紹介することとする．

- 高橋・鈴木 (2005) プロ野球チームについて，ファンになった理由とチームへの提言を自由記述，数量化 III 類を実施し，プロ野球チームに対する愛着心や満足度にどのような要因がかかわっているのか，ファンが球団に対して何を望んでいるのかを確認している．
- 磯島 (2006) や磯島 (2010) では，農作物の品質や価格に関する自由記述アンケートから，頻出単語の確認や，購入時に重視する商品属性とキーワードとのコレスポンデンス分析，キーワードの等質性分析により，購入者のニーズ把握を行なっている．
- 大瀬良他 (2007) 大瀬良 (2008) は，通信販売会社の化粧品カテゴリーに電話やインターネットで寄せられた顧客の声の分析．声の内容分類，発言内容と発言前後の顧客ロイヤルティの変化について調べている．声を上げた顧客のロイヤルティは向上することがわかった．数量化 III 類とクラスター分析の結果から声を 4 分類を行い，声のカテゴリごとの，購買継続期間の分析も実施している．
- 三川他 (2007) は，ある商品を買い続ける顧客にその理由をアンケート，自由記述データに基づいて顧客ロイヤルティの構造を視覚化，数量化 III 類を用いて高いロイヤルティを持つ顧客の特徴抽出をしている．
- 伊藤 (2007) は，美術館来訪者に対して，生活における美術館の位置づけ，美術鑑賞はどんな意味を持つのかという自由記述のアンケート．その結果と属性とのコレスポンデンス分析を行っている．また，主用な語との係り受けの分析もある．
- 小代 (2008) は，浴室の好みについての自

由記述のネットワーク分析を行っている．

- 菊池 (2008) は，おたくのイメージを表す単語と自分はおたくという認識があるかのコレスポンデンス分析を行っている．
- 浅川・岡野 (2009) は，飲料の CM を女子大学生に見せ，登場するタレントの好き嫌いと，好きな理由の自由記述から，係り受けの頻度確認を実施している．
- 伊藤・曽和 (2010) では，yahoo の blog 検索を用いて，場所名＆庭で検索，そのデータを分析している．庭園と単語のコレスポンデンス分析を行っている．
- 川島他 (2010) では，ゲーム批評雑誌に載っているゲーム批評記事を分析対象とし，単語の共起頻度を元にしたクラスタリングを行い, 年代ごとの単語クラスターの出現頻度の変化を集計している．
- 庄司 (2010) では，ある自分が好んでいる店舗について，その店舗を推奨できる理由を自由記述させ，同時にその店舗に対するロイヤルティの高さも測定し， ロイヤルティの高さと出て来やすい単語との関係性を分析している．たとえば，ロイヤルティの高いグループでは推奨の理由として「豊富」などの単語が出やすいことが確認された．
- 森脇他 (2010) では，学食で提供したヘルシー定食について，それが好まれる理由を自由記述させ，その回答から出てきた単語の頻度を確認している．
- 粕渕・松村 (2011) は大学生協のひとことカードとそれに対する返答の分析である．投稿意図の分類を行い，投稿意図の中で特に「要望」に分類されるものについて，返答を実現段階 (実現，不可能，検討，努力，問い直し) に分類している．また，ポジティブ内容は実現段階の高いテキストに多いなど，返答に使われた語句と実現度の関係性の確認，表現内容の確認なども実施している．

企業研究への応用例も数多くある．喜田 (2006) では，アサヒビールの有価証券報告書から，名詞の数の時系列的な変化とシェアや利益などの変化との関連性 (相関) の分析を行い，また，主要な概念について，時系列的に内容がどのように変化しているのかを調べている．白田他 (2009) も有価証券報告書を分析した例である．同論文では，倒産企業と継続企業の特徴を明らかにするために，倒産企業に特徴的な語，継続企業に特徴的な語などを確認している．小田・三橋 (2010) では，製造業 121 社の経営理念を分析対象とし，企業のクラスター分析，語をまとめて企業クラスターごとの使用頻度の分析，各クラスターのパフォーマンスの分散分析などを行っている．

記虎 (2009a,b, 2010b,a) では，企業の CSR 基本方針について取り上げている．主な分析内容は，係り受けの分析や，その結果からの企業のクラスタリングなどである．

これら以外にも，滝岸・町田 (2007) のように，農家自身に収支結果の現状，原因，解決方法を自由記述させ，その結果について係り受けの分析を行い，原因と解決方法のクロス集計を実施した例もある．

医歯薬看護学

経営学ほどではないにせよ，医歯薬看護学もテキストマイニングが頻繁に応用されている分野である．その応用は，2006 年以降に集中している点が注目に値する．

医療，それから後述する工学に共通した応用として，インシデントやエラーの内容分析が挙げられる．岡部他 (2006) では，インシデントに関して, 発生場所や時間などのメタデータとテキストを用いた共起情報のネットワーク分析が行われている．五十嵐・福士 (2011), 五十嵐他 (2010) でも，放射線技師に対して，経験したエラーの内容記述をしてもらい，原因について頻出単語の確認し，また，クラスター分析でエラー発生状況のデンドログラムを作成し，同時に発生しやすい状況を示す単

語を確認している．

医療に関わるアンケート調査における応用例も見られる．村上他 (2009) では，病院における体位変換に使う用具についてのアンケートを行い，自由記述に出てきた単語の頻度や係り受けの分析，語のマッピング，因果関係のネットワーク作成を行っている．七海他 (2011) は，ケアマネジャーから薬剤師や薬局に対する意見のテキストマイニングを行ったもので，単語の出現頻度や頻度や共起分析を行い，ケアマネージャーのニーズを探っている．

また，二見他 (2010) のように，胸部 CT 検査の検査レポートの内容を分析，類似記載を特定する応用例もある．

工学

工学分野における応用事例は，先述の通りエラー分析にかかわるものが多く見られる．安藤他 (2002)，安藤・大和 (2004) では，船舶の故障報告書について，発生した故障の内容の抽出や，出現単語の組み合わせを元にしたトラブルの発生頻度の算出，1 件あたりの平均遅延時間の算出を行っている．

同様に，北澤・長田 (2008) では，自動車のリコール情報を分析し，不具合部位や一次要因についてキーワード抽出を行い，企業ごとの頻度の比較や，不具合状況の重大度でレベル分けした時系列による比較を実施している．西浦・山田 (2010) においても自動車のリコール情報が分析対象となっている．同論文では，不具合に関する用語の頻度や共起ネットワーク，部品と現象との間の数量化 III 類，部品間の関係に関するアソシエーションルールといった分析が行われている．

野守他 (2010) では，子供の傷害データを用いて，どのような製品に対してどのような行動が行われているのか分析している．具体的には，製品の種類の頻度分析，個々の製品に対する行為の頻度分析．年齢，製品，行動，事故内容それぞれの関連性のベイジアンネットワークによる分析である．

エラー分析に関わるもの以外では，尾暮他 (2004) の，脱原発を主張するコミュニティの概念体系と技術者コミュニティの概念体系との比較を行った事例が挙げられる．同論文では，核燃料リサイクルに関する解説記事と，脱原発を主張するコミュニティのサイトそれぞれのデータについて，自己組織化マップを用いて分析し．その結果から，両者の知識や意見の効率的な共有を目指している．

経済学

経済学におけるテキストマイニングは，景気動向，経済動向に関連するテキストから，実際に景気，市場の動向を予測，説明するということがメインテーマとなっている．和泉他 (2007) では人工市場に現実のニュースを導入することでより現実に即したシミュレーションを行うため，ニュースのテキストマイニングを行なっている．テキストの特徴を決定木で分類することで，テキストデータから自動的に経済動向を推定するという内容だ．同様に，和泉他 (2010)，和泉他 (2011) では日銀の金融経済月報を利用し，単語の共起ネットワークを作成，また，主成分分析による単語のグループ化を行い，その主成分スコアから 2 週間後の市場価格の予想を行う回帰分析を実行している．谷口他 (2011) では，SVM を用いて新聞経済記事の分析を行い，段落が経済動向をネガティブに捉えているか，ポジティブに捉えているか，その他かに分類している．

それ以外の応用事例としては，有村・坂本 (2002) で，経済関係のニュースのうち海運に関する記事とその他の記事から，海運関係の記事に特徴的な内容を示している例がある．

心理学

心理学では，被験者の自由記述に対する分析に応用されるケースが多い．アンケートの自由記述項目に対する分析は特に多い．真船他 (2006) では，ストレッサーを自由記述させた上でキーワードの頻度を分析，数量化 III 類

で分析した結果を元にした回答対象者のクラスタリングを行っている．川島他 (2009) では，自殺を希望した患者への医師のメッセージ内容分析対象としている．同論文では，性別や年齢とのコレスポンデンス分析．そのスコアを用いた単語のクラスター分析などが行われている．KUSUMI et al. (2010) は，ノスタルジアを感じるシチュエーションに関する自由回答について，単語のクラスター分析などを行った事例である．緒方他 (2010) においては，司法解剖に関する遺族へのアンケートの自由記述欄について，MDS やクラスター分析による分析を試みている．

実験に対する自由回答への分析も見られる．岡本他 (2008) では，心理学実験としてのゲーム参加者に，ゲーム内で用いた地域について，その印象を記述させたものを用いて地域とコレスポンデンス分析やクラスター分析を行っている．岡本他 (2009) は，大学での一週間を写真に撮影させ，その説明を分析対象としている．ここでは，語句と大学とのコレスポンデンス分析，語句のクラスター分析が行われている．

安田・鳥山 (2007) は，これらの例とはやや異なる．同論文では，電子メールによるコミュニケーションの内容が分析対象となっている．受信者と送信者の関係とコミュニケーションの内容の分析や，企業におけるパフォーマンスの高い層のコミュニケーションでは，ポジティブな内容が多いといった，特徴の抽出が行われている．

教育学

教育分野における応用は，自動採点の研究など入試に関わるもの，授業に関わる学生へのアンケートの分析などがあり，どちらも学生からの働きかけに対して適切なフィードバックを返すための研究と言える．

入試に関わるものとしてはまず小論文の自動採点が挙げられ，国内におけるその代表的研究は，石岡 恒憲・石岡 恒憲 (2003) であろう．アメリカにおいて用いられている小論文採点システム e-rater を参考に，文体，論理構成，内容の観点から日本語小論文を自動的に採点するシステムを作成，公開している．自動採点以外では，吉村 (2009) にの，AO 入試の選考書類の分析が挙げられる．この論文では，教育学部の AO 入試の選考書類について，内容のクラスタリングを行い，志望理由と希望校種の連関など，内容と属性の関連性を調べている．

授業に関わる学生からの働きかけの分析としては，佐川他 (2004) が挙げられる．ここでは，授業で作成した看護研究抄録を元にして，共起ネットワークの分析などが試みられている．同様に，濃沼他 (2008) では，薬剤師の実務実習に先立ち，自由記述のアンケートで学生がどのような分野に関心をもつのか，頻度分析やコレスポンデンス分析により検討を行っている．谷塚・東原 (2009) では，現場での実習科目の感想等について，単語間のクラスタリングや，行き先と単語とのコレスポンデンス分析などによる分析を行っている．さらに，鈴木他 (2009) でも，薬学部学生の実務実習を受け入れた病院の実習指導者を対象に，事前学習に対する期待の理由と事前学習に対する印象について自由記述させ， 病院における職種とキーワードのコレスポンデンス分析や，「必要」や「不安」といった特性の原因となる単語の抽出を行っている．高橋他 (2009) では，中学生に対して，ストレッチはどのようなものと認識しているか，その内容について確認している．

若干異なる例としては，椿他 (2010) がある．ここでは，学習の改善のために，学生に PDCA，CAPD サイクルを割り当て，その学習プロセスを記録させたものを分析している．両サイクルにおける文字数のカウント，出てくる単語を用いたコレスポンデンス分析，その結果を利用したクラスター分析である．

文学

序論においても述べたように，文学作品における著者推定の問題は，テキストマイニングという言葉が生まれる遙か昔から存在し，これらをテキストマイニングという理論を応用した事例として紹介するのは不適切とも言える．また，著者推定を行っている論文も極めて多数存在するため，それらを紹介するのは困難である．そこで，ここでは現在この分野の中心を担っていると思われる二氏の論文を紹介するにとどめる．

確認できた限り，1970 年代には既に，村上征勝氏による文学作品の著者推定の論文が存在する．たとえば村上 征勝 (2002) では，日蓮の著作とされているもののうち，真贋が不明である書物に対して，単語の出現率を用いたクラスター分析による分類からの推定を試みている．また，同論文では，源氏物語のうち，他者の著作であるという説のある宇治十帖に関して，特定の名詞の出現率の確認，頻出助動詞を用いた数量化 III 類からの推定も試みている．

この問題については，金明哲氏の論文も多い．金・村上 (2007) では，10 人の作家の小説と，6 人の書き手の日記について，1 編につき平均 1 回以上出てくる単語に絞った上で， その単語を元にしてランダムフォレスト法により著者の推定を行っている．

法・政治学

法律文書はある程度定型化しているため，分析対象として有用であるように見えるが，現状では判例の分析と特許文書の分析以外応用事例は発見できなかった．特許文書の分析については，学術研究への応用事例として後述する．川島他 (2010) では，判例データベースを利用した知識マップを作成している．

政策研究においては，答申書の分析が見られる．崔・浅見 (2004) では，住宅建設五箇年計画の答申，計画について，各期 (5 年× 8 期) において，それぞれどのような単語が頻出であったのか，期と単語とのコレスポンデンス分析による期ごとの特徴の確認を行っている．そして，答申と計画との類似度の分析，共起タームのネットワーク分析なども実施されている．また，佐藤他 (2011) のように，豪雨に対する教訓，課題等の自由記述データと，自治体が経験した雨量と当日の雨量のデータを利用して，豪雨を経験したことのある市町村としたことのない市町村，降水量の大きかった市町村，少なかった市町村それぞれに特徴的なキーワードの抽出，キーワードのクラスター分析を行った事例もある．

学術

テキストマイニングは，学術研究の基礎として使われることも多い．論文や特許データベースの内容分析から，研究内容，課題の把握などが可能だからである．

景山・辻 (2005) では，大学のウェブサイトを分析対象とし，経営工学系の研究内容について，TF/IDF 値から大学の特徴を抽出している．類似した研究として篠原他 (2007), Masanori et al. (2008) がある．ここでは，特にアンカーテキストに注目し，研究室情報に特徴的な文字列の自動抽出を行っている．

論文の内容分析は，医学，工学分野に多く見られる．佐々木他 (2005) では，正規表現を利用して，論文アブストラクトから原子状態を抽出している．小池 (2007) では，医学生物学分野における論文データベースのテキストマイニングに用いられる BioTermNet 開発し，同ソフトを用いて，概念ネットワークの作成．遺伝子と機能，疾患との関係などを調べる例を示している．

この分野において，最も応用が進んでいるのは，企業ニーズも高い特許文書の分析であろう．豊田・菰田 (2011) のような書籍が出版されていることからも，関心の高さが窺える．石川他 (2004) では，繊維工学の分野で，特許文献から化合物とその性能との因果関係を抽出，整理している．酒井他 (2009) では，特許

情報から技術課題情報を取り出すための手がかりとなる条件を調べ，実際に技術課題情報を抽出する例を示している．

論文と特許情報の両方を分析対象としている例もある．山本 (2009) では，論文，特許情報に出てくる単語の類似性マップを作成し，企業と論文の著者の類似性を確認している．また，落合他 (2010) では，特許情報，科学論文に関するデータベースを作成し，それを用いた特許情報に出てくる単語のマッピングの例を示している．

考察・展望

これまでに説明してきた応用事例を，分析手法に着目して整理すると，概ね以下の通りに分類できる．

1. 単語の出現頻度の集計
2. 係り受けの頻度の集計
3. SOM，MDS による単語のマッピング
4. ベイジアンネット等による単語のネットワーク分析
5. コレスポンデンス分析，数量化 III 類による単語と属性，対象の同時布置
6. 対象のクラスタリング
7. SVM 等を用いたテキストの分類
8. キーワードの自動的な抽出

確かにテキストマイニングには多くの応用事例がある．しかし，分析という観点から見ると，用いられている手法はかなり絞られている．単純な集計を行うか，単語間の同時出現の割合を分析するか，テキストの属性の特徴を出現単語を用いて分析するといった，記述的な分析手法が大半を占めている．推測的な手法としては，機械学習によるテキストの分類が主となっている．

テキストデータの場合，形態素解析され，また，係り受け解析されたデータから必要な部分をいかに抽出して実際に用いるデータとするかは，本来的には分析者に委ねられているものであり，研究の個性も出る部分であるが，たとえば小論文の自動採点システム，文学作品の著者推定の問題のように，個々の目的に合わせたデータの作り方をするのは応用研究を行うものには困難が大きい．ソフトウェアが示す手順通りに分析を行うとなると，基本的な名詞を抽出して単語間の関連を見る，単語と属性の関係を見るという段階に留めざるを得ないのが現状なのだろうと思われる．また，一般的なデータマイニングと同様，統計的仮説検定等の基本的な推測統計の手法を用いることが難しい点も，応用事例が限られている理由であろう．

この問題を解決するためには，より多彩な分析手法を簡単な操作で実行可能なソフトウェアが必要であると同時に，応用研究者がテキストデータ分析の可能性をより強く感じ，こういった分析をしたいと考えられる状況が必要となるのだろう．

参考文献

安藤英幸・大和裕幸・堀晃・増田宏・白山晋 (2002). テキストマイニングを用いた故障報告書分析手法の研究　日本造船学会論文集，**2002**(192)，475–483.

安藤英幸・大和裕幸 (2004). テキストマイニングによる船舶故障データの分析 (〈 特集 〉製造現場における信頼性)　日本信頼性学会誌：信頼性，**26**(8)，906–912.

有村博紀 (2003). テキストマイニング：ウェブデータからの知識発見を目指して　日本化学会情報化学部会誌，**21**(2)，28.

有村博紀・坂本比呂志 (2002). テキストマイニングにおける最適パターン発見 (〈 特集 〉データ・テキストマイニング)　応用数理，**12**(4)，366–378.

浅川雅美・岡野雅雄 (2009). テレビ CM に登場するタレントに対する態度を決定する要因の分析–自由記述のテキスト・マイニング　広告科学，**50**，91–98.

崔延敏・浅見泰司 (2004). 言語統計分析による住宅建設五箇年計画及び答申の特性分析: 政策の立案と評価における非定型・大量情報の活用可能性　日本建築学会計画系論文集 (579)，89–96.

二見光・山岸宏匡・川口修・塚本信宏・藤井博史・笠松智孝・安藤裕・長田雅和・久保敦司 (2010). 構造化技術を用いた読影レポートの類似記載を特定する手法の開発　日本放射線技術學會雑誌，**66**(9)，1229–1236.

長谷川久 (2011). 特集号投稿論文テキストマイニングの利用による早期人材育成の実践 –コール・ログ分析による要員育成の効率化 (特集 コンタクトセンタ)　情報処理学会デジタルプラクティス，**2**(3)，192–199.

Hearst, Marti A. (1999). Untangling text data mining in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* -, 3–10, Morristown, NJ, USA: Association for Computational Linguistics.

樋口耕一 (2004). テキスト型データの計量的分析：2 つのアプローチの峻別と統合　理論と方法，**19**(1)，101–115.

市村由美・鈴木優・酢山明弘・折原良平・中山康子 (2003). 日報分析システムと分析用知識記述支援ツールの開発　電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理，**86**(2)，310–323.

五十嵐博・福士政広・星野修平 (2010). テキストマイニングを用いた診療放射線技師のヒューマンエラー分析　日本保健科学学会誌，**13**(2)，59–70.

五十嵐博・福士政広 (2011). 質問紙票を用いた放射線治療における診療放射線技師のヒューマンエラー分析　日本保健科学学会誌，**14**(1)，40–48.

石田基広 (2008).　R によるテキストマイニング入門 [単行本（ソフトカバー）]，森北出版.

石川大介・石塚英弘・宇陀則彦・藤原譲 (2004). 特許文献における因果関係の抽出と統合情報知識学会誌，**14**(4)，105–118.

石川修・星野敏 (2004). テキストマイニングを用いた都市農村交流ニーズの把握 : 岡山県吉永町ふるさと村の八塔寺山荘の落書き帳を対象として　農村計画学会誌，**23**，181–186.

石岡 恒憲・亀田 雅之 (2003). コンピュータによる小論文の自動採点システム Jess の試作計算機統計学，**16**(1)，3–19.

磯島昭代 (2006). テキストマイニングを用いた米に関する消費者アンケートの解析　農業情報研究，**15**(1)，49–60.

磯島昭代 (2010). テキストマイニングによる農産物に対する消費者ニーズの把握　フードシステム研究，**16**(4)，4_38–4_42.

伊藤大介 (2007). テキストマイニング手法を用いて分析した美術館来館者の生活における美術館の存在意義: 静岡県立美術館来館者アンケートを事例として　文化経済学，**5**(3)，101–110.

伊藤いずみ・曽和治好 (2010). ブログからみる日本庭園の評価　ランドスケープ研究，**73**(5)，377–380.

和泉潔・松井宏樹・松尾豊 (2007). 人工市場とテキストマイニングの融合による市場分析　人工知能学会論文誌，**22**，397–404.

和泉潔・後藤卓・松井藤五郎 (2010). テキスト情報による金融市場変動の要因分析　人工知能学会論文誌，**25**(3)，383–387.

和泉潔・後藤卓・松井藤五郎 (2011). テキスト分析による金融取引の実評価　人工知能学会論文誌，**26**(2)，313–317.

金明哲 (2009a).　テキストデータの統計科学入門 [単行本]，岩波書店.

金明哲 (2009b). 文章の執筆時期の推定 : 芥川龍之介の作品を例として　行動計量学，**36**(2)，89–103.

金明哲・村上征勝 (2007). ランダムフォレスト法による文章の書き手の同定 (特集文化を科学する)　統計数理，**55**(2)，255–268.
景山明宣・辻洋 (2005). TF/IDF アルゴリズムを用いた研究機関の特徴抽出法　電気学会論文誌. C, 電子・情報・システム部門誌，**125**(5)，713–719.
粕渕孝文・松村真宏 (2011). サービス利用者の要望に含まれる語句とその実現率との関係　経営情報学会誌，**19**(4)，385–393.
川島大輔・小山達也・川野健治・伊藤弘人 (2009). 希死念慮者へのメッセージにみる，自殺予防に対する医師の説明モデル：―テキストマイニングによる分析　パーソナリティ研究，**17**(2)，121–132.
川島啓・ロベルアダム・山田健智・大竹裕之 (2010). 社会的ニーズを踏まえた法律情報に対する知識構造マップの開発　情報知識学会誌，**20**(2)，207–214.
川島隆徳・村井源・往住彰文 (2010). ゲーム批評から見たゲームの「面白さ」–レビューテキストの計量解析による叙述対象の自動抽出 (特集 ゲームのユーザエクスペリエンス研究)　デジタルゲーム学研究，**4**(1)，69–80.
喜田昌樹 (2006). アサヒの組織革新の認知的研究–有価証券報告書のテキストマイニング　組織科学，**39**(4)，79–92.
菊池聡 (2008). 「おたく」ステレオタイプの変遷と秋葉原ブランド (特集地域ブランディングの原点)　地域ブランド研究 (4)，47–78.
北澤謙・長田洋 (2008). 公開情報に基づく品質事故の分析手法の提案とその成果：自動車におけるリコール分析　品質，**38**(1)，147–155.
記虎優子 (2009a). 企業の社会的責任 (CSR) に対する基本方針による企業の類型化–テキストマイニングによるクラスター化の試み　社会情報学研究，**13**(1)，17–29.
記虎優子 (2009b). 企業の社会的責任 (CSR) の一環としての情報開示志向と企業ウェブサイトにおける情報開示の関係–テキストマイニングを利用して　会計プログレス (10)，28–42.
記虎優子 (2010a). CSR 基本方針に表れた企業の環境志向と EMS 構築度の関係　環境技術，**39**(8)，486–492.
記虎優子 (2010b). 企業のステークホルダー志向と情報開示の関係：企業ウェブサイトに着目して　環境技術，**39**(2)，103–111.
小池麻子 (2007). 3 テキストマイニングによる潜在的知識の発見支援 (〈 特集 〉情報の価値化・知識化技術の実現へ向けて)　情報処理，**48**(8)，824–829.
濃沼政美・小池勝也・中村均 (2008). 実務実習事前教育に向けたテキストマイニング手法の活用　藥學雜誌，**128**(6)，925–931.
工藤拓・松本裕治 (2001). チャンキングの段階適用による係り受け解析　情報処理学会研究報告. 情報学基礎研究会報告，**2001**(20)，97–104.
黒橋禎夫 (2000). 結構やるな, KNP(〈 特集 〉使いやすくなった自然言語処理のフリーソフト：知っておきたいツールの中身)　情報処理，**41**(11)，1215–1220.
黒岩祥太 (2005). ブランドイメージと消費者接点の関連についてのテキストマイニングマ-ケティングジャ-ナル，**25**(1)，38–50.
KUSUMI, TAKASHI, KEN MATSUDA, & ERIKO SUGIMORI (2010). The effects of aging on nostalgia in consumers' advertisement processing *Japanese Psychological Research*, **52**(3), 150–162.
真船浩介・鈴木綾子・大塚泰正 (2006). 大学生におけるストレッサーの特徴–認知的評定、及び心理的ストレス反応との関連の検討　学校メンタルヘルス，**9**，57–63.
Masanori, SHINOHARA, CHIKURA Shinsaku, & HADA Yoshiaki (2008). Auto-

matic Extraction of Academic Research Information from Higher Education Institution Websites Using Anchor Texts and Link Structures *Educational technology research*, **31**(1), 143–151.

松本裕治 (2000). 形態素解析システム「茶筌」(〈 特集 〉使いやすくなった自然言語処理のフリーソフト: 知っておきたいツールの中身)　情報処理，**41**(11)，1208–1214.

松村真宏・三浦麻子 (2009).　人文・社会科学のためのテキストマイニング [単行本]，誠信書房.

松村真宏 (2008). テキストデータのマーケティングへの活用と課題　経営システム = Management systems : a journal of Japan Industrial Management Association，**18**(1)，32–37.

三川健太・高橋勉・後藤正幸 (2007). テキストデータに基づく顧客ロイヤルティの構造分析手法に関する一考察　日本経営工学会論文誌，**58**(3)，182–192.

森脇弘子・山崎初枝・前大道教子 (2010). 学生食堂におけるヘルシー定食提供の試み　日本調理科学会誌，**43**(6)，359–365.

村上亜紀・滝沢茂男・木村哲彦・長岡健太郎・森田能子 (2009). 褥瘡予防における福祉用具の役割とその利用の実際の研究　バイオフィリアリハビリテーション研究，**5**(1)，1–10.

村上 征勝 (2002). 2. 年代・産地・個人推定：2-3 著者を探る古文書の計量分析 (〈 特集 〉いにしえの世界を探る科学技術)　電子情報通信学会誌，**85**(3)，158–161.

七海陽子・恩田光子・櫻井秀彦・田中理恵・坪田賢一・的場俊哉・向井裕亮・荒川行生・早瀬幸俊 (2011). 在宅ケアにおける薬剤師業務に対するケアマネージャーの情報収集手段及び意識・要望に関する調査研究 *YAKUGAKU ZASSHI*, **131**(5), 843–851.

那須川哲哉 (2001). コールセンターにおけるテキストマイニング (〈 特集 〉「テキストマイニング」)　人工知能学会誌，**16**(2)，219–225.

那須川哲哉 (2009). テキストマイニングの普及に向けて：研究を実用化につなぐ課題への取組み　人工知能学会誌，**24**(2)，275–282.

那須川哲哉・諸橋正幸・長野徹 (1999). 2 テキストマイニング：膨大な文書データの自動分析による知識発見 (〈 特集 〉フィールドを広げる自然言語処理)　情報処理，**40**(4)，358–364.

西浦友子・山田秀 (2010). 不具合情報に基づくデザインレビュー項目構築に関する研究品質，**40**(4)，411–419.

野守耕爾・北村光司・本村陽一・西田佳史・山中龍宏・小松原明哲 (2010). 大規模傷害テキストデータに基づいた製品に対する行動と事故の関係モデルの構築：エビデンスベースド・リスクアセスメントの実現に向けて人工知能学会論文誌，**25**(5)，602–612.

落合圭・小林義英・橋本定幸・塩尻栄美子・山崎雅和・栗原正昭・浜中寿・坂内悟・國谷実・治部眞里 (2010). サイエンスリンケージによる JST 事業成果分析（下）可視化の具体的手法　情報管理，**52**(11)，651–659.

小田恵美子・三橋平 (2010). 経営理念と企業パフォーマンス–テキスト・マイニングを用いた実証研究 (特集 CSR、企業倫理、企業理念は本当に役に立つのか)　経営哲学，**7**(2)，22–37.

緒方康介・西由布子・前田均 (2010). 犯罪・事故等関連死亡者の遺族における司法解剖への想い–自由記述文に対するテキスト・マイニングを用いた分析　犯罪学雑誌，**76**(2)，41–47.

小木しのぶ (2005). ことばによる感性と映画–テキストマイニングによる感性の抽出 (エンタテインメント感性特集)　感性工学研究論文集，**5**(3)，43–47.

尾暮拓也・高松悠・古田一雄 (2004). コミュニティを超えた知識共有のための原子力安全オントロジー設計方法　社会技術研究論文集，**2**，389–398.

岡部貴博・吉川大弘・古橋武 (2006). メタデータと語句の共起情報を利用したインシデントレポート解析システムの提案 (〈 特集 〉テキストの可視化と要約)　知能と情報：日本知能情報ファジィ学会誌, **18**(5), 689–700.

岡本青史・関口実・三末和男・西野文人 (2000). カスタマーセンター支援システム　人工知能学会誌，**15**(6)，1027–1034.

岡本卓也・藤原武弘・野波寛・加藤潤三 (2008). 共有集団イメージ法を用いた集団間関係の解析の試み　実験社会心理学研究，**48**(1)，1–16.

岡本卓也・林幸史・藤原武弘 (2009). 写真投影法による所属大学の社会的アイデンティティの測定　行動計量学，**36**(1)，1–14.

大瀬良伸 (2008). 顧客の声と購買行動の関係性　商品研究，**55**(3)，57–68.

大瀬良伸・中野香織・松本大吾 (2007). 声の発生に伴う顧客ロイヤルティの変化について *Direct marketing review*，**6**，21–42.

佐川輝高・岡田ルリ子・青木光子 (2004). 学生の看護研究抄録におけるテキストマイニング法の検討　看護と情報：看護図書館協議会会誌，**11**，36–41.

酒井浩之・野中尋史・増山繁 (2009). 特許明細書からの技術課題情報の抽出　人工知能学会論文誌，**24**(6)，531–540.

櫻井茂明・酢山明弘 (2002). ファジィ帰納学習におけるキー概念集合を含む属性値の扱い　日本ファジィ学会誌, **14**(6), 640–647.

櫻井茂明・酢山明弘 (2005). キーフレーズに基づいたテキストの分析 (〈 特集 〉理解技術におけるソフトコンピューティング)　知能と情報：日本知能情報ファジィ学会誌，**17**(1)，52–59.

佐々木明・村田真樹・金丸敏幸・白土保・井佐原均・上島豊・山極満 (2005). 論文アブストラクトから原子分子の状態の情報を検出，抽出する方法の研究　プラズマ・核融合学会誌，**81**(9)，717–722.

佐藤翔輔・林春男・田村圭子・浦田康幸 (2011). 平成 21 年の大雨時の避難勧告発令経験にもとづく自治体の対応に関する教訓・課題–大雨災害における避難のあり方等検討会「避難勧告・避難指示を発令した市町村に対する調査」の自由回答の TRENDREADER(TR) 解析　自然災害科学，**30**(1)，123–145.

篠原正典・地蔵真作・葉田善章 (2007). リンク情報を基にした高等教育機関 Web からの研究室情報の自動抽出 (〈 特集 〉学習オブジェクト・学習データの活用と集約)　日本教育工学会論文誌，**31**(3)，383–391.

白田佳子・竹内広宜・荻野紫穂・渡辺日出雄 (2009). テキストマイニング技術を用いた企業評価分析：倒産企業の実証分析　年報経営分析研究 (25)，40–47.

小代禎彦 (2008). 個人差を考慮した浴室の好みの評価 (特集 [日本感性工学会] 第 9 回大会)　感性工学，**8**(1)，53–60.

庄司真人 (2010). 顧客ロイヤルティと推奨の関係　日本経営診断学会論集, **9**, 103–108.

鈴木慎一郎・濃沼政美・日高由加里・小池勝也・中村均 (2009). 実務実習事前学習に対する実務実習受け入れ側の意識調査と解析 — 日本大学薬学部における取り組み — *YAKUGAKU ZASSHI*, **129**(9), 1103–1112.

高橋亮輔・林英俊・渋川正人・中村崇・掛川晃・関賢一・武藤芳照 (2009). 中学生のストレッチングの実施状況および認識度について：ースポーツ障害との関連ー　身体教育医学研究，**10**(1)，43–49.

高橋大地・鈴木秀男 (2005). プロ野球チームに対するロイヤルティと満足度に関する研究　品質，**35**(1)，139–145.

滝岸誠一・町田武美 (2007). テキストマイニン

グシステムをもちいた経営方針意思決定手法の研究　農業情報研究, **16**(3), 113–123.
谷口将太・坂地泰紀・酒井浩之・増山繁 (2011). 経済新聞記事から抽出した景気動向を示す根拠表現への極性付与手法の提案 (研究速報)　電子情報通信学会論文誌. D, 情報・システム，**94**(6)，1039–1043.
豊田裕貴・菰田文男 (2011).　特許情報のテキストマイニング–技術経営のパラダイム転換，ミネルヴァ書房.
豊田裕貴・森永聡 (2003). 企業におけるマーケティング分野でのテキスト活用事例：ブランド・イメージ調査へのテキストマイニング技術の適用 (自然言語処理技術による情報マネジメントの実際)(〈 特集 〉自然言語処理の高度化による知的生産性の向上) 情報処理，**44**(10)，1028–1031.
椿美智子・小林高広・久保田一樹 (2010). 学習型 PDCA 及び CAPD サイクルを用いた学習過程テキスト情報の個人差を考慮した分析　教育情報研究：日本教育情報学会学会誌，**25**(4)，15–27.
上田芳弘・成田仁志・加藤直孝・林克明・南保英孝・木村春彦 (2004). テキストマイニングと強化学習を用いた電子メール自動分配 (データマイニング)　電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, **87**(10)，887–898.
山本外茂男 (2009). 産学連携のマッチング性分析におけるテキストマイニングの有効性 情報の科学と技術，**59**(6)，291–297.
安田雪・鳥山正博 (2007). 電子メールログからの企業内コミュニケーション構造の抽出 (特集 ソーシャル・キャピタルの組織論) 組織科学，**40**(3)，18–32.
谷塚光典・東原義訓 (2009). 教員養成初期段階の学生のティーチング・ポートフォリオのテキストマイニング分析: INTASC 観点「コミュニケーション」に関するリフレクションの記述から　日本教育工学会論文誌，**33**，153–156.
吉村宰 (2009). AO 入試選考書類のテキストマイニング　大学入試研究ジャーナル (19)，157–160.