

《論 文》

# 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

山 本 真 照

## 1. はじめに

本稿の目的は、企業実務における実践的な知識の活用方法についての最新の取り組み事例をサーベイして、実践知識活用の諸条件について考察するとともに、実践知識活用の新しい方法、とりわけテキストマイニング手法の洗練に向けた将来課題を抽出することである。

実践知識に注目した背景には近年の知識経済化の進展がある。近年では、組織に埋め込まれ、文脈に依存して、他社が容易に模倣できない知識を不断に作り上げ、活用できる企業のみが市場競争力を維持することができる。しかしこれは多くの企業にとって実務的に困難な課題である。とりわけ知識を経営のあらゆるプロセスにおいて活用することは、多くの企業で最重要課題となっている。科学技術の知識の重要性はとりわけ製造業などのモノ作り企業においては疑うべくもないが、今日ではこうした比較的体系だった知識とは異なる、実践的な知識の有用性が増している。顧客や市場の知識の取り込み、蓄積、共有は、あらゆる企業経営にとっての現代的課題であろう。

知識の取り込み、蓄積、共有を含む知識活用の実践においては、ともすれば永久に忘れられてしまうような「普通の人々の日常的な業務遂行」に光を当てることにこそ本質があるのではないかと考える。重要な科学的発見や、政治力を駆使した成功談、などは容易に伝搬するが、それがすぐに日常の業務に役立つとは限らない。知識の粒度が違うからである。これまでの研究では比較的大き

な粒度の知識活用が扱われてきたようであるが、企業の継続的なイノベーションの主役は、あくまでも日常的な改善活動の中に存すると筆者は考える。本稿はこの仮説に対して具体的な検証を加えていることが、従来研究に対する特徴であると考えている。

企業において知識を活用することの重要性自体はこれまで様々に論じられてきたが、その方法についての決定的な解は示されておらず、常に課題として認識されている。こうした課題に対応する経営手法は、ナレッジマネジメントと呼ばれ、経営学の一分野として発展を続けている。ナレッジマネジメントでは、企業に散在する情報を結集し、後に再利用できるような形式に変換することが、企業全体の知識共有を促進することであり、最も重要な活動の一つであると考えている。それでは、この活動を促進するための具体的な方法は何であろうか。

それに対しては、データマイニングやテキストマイニングなどの統計的手法、とりわけコンピュータリソースの利用を前提としたこれらの手法の適用が、非常に有効となる。なぜなら、企業には膨大な量の情報が保管されており、これらは多くの場合、価値に関して玉石混淆であって、コンピュータを用いた大量処理によって一定の基準で処理をさせることは、人間の情報処理能力を助ける非常に大切な手段だからである。

本稿の分析にテキストマイニングを用いる別の側面としては、これが日常業務で発生する膨大な情報を処理するための有用なツールである、ということが挙げられる。たとえば、業務日報などを

思い浮かべてみよう。業務日報は担当から課長、課長から部長、部長から事業部長へとそのレポート階層があがるにつれ、現場の知が抽象化され、かつ、上層部に都合のよいストーリーに変化してしまうということ、よく起こり得ることである。このため、現場レベルでの最良の方針を検討するためには、オリジナルデータにさかのぼることが一番の方法であろう。しかし、すべての日報をすべての人が参照するには膨大すぎて、人間の情報処理能力の限界を超えてしまうことが多い。そのような場合に、適切な人に適切な知を伝達するには、テキストマイニングのような機械処理のツールが有用である。

第1項では本稿の問題意識と目的を述べた。以下、第2項では、本稿の問題意識の背景の説明と先行研究のレビューを行い、第3項では本稿の目指すナレッジマネジメントの新たな局面についての仮説を提示して、実際にテキストマイニング分析によって仮説の妥当性を検証する。第4項では、検証で得られた結果をまとめ、本稿の課題と今後の展望を述べる。

## 2. 問題設定の背景と先行研究のレビュー

現代の企業活動においては、大量の情報を処理することが必須である。そして、膨大になりすぎた情報の海に立ち向かうにあたり、人間の認知能力の限界を補うために情報システムを利用することは不可欠な要件となっている。当たり前であるが、こうした事実は、企業のコスト増にもつながっている。しかし、このコストを上回るメリットを追求し、企業の競争力を高めることは可能である。現に、情報システムの使い方の巧拙によって、企業間で競争力の差異が生まれている。この差異はどうして生じるのだろうか。情報システムをつかいこなす企業と、十分につかいこなせていない企業では、その使い方によどのような特徴的な差異があるのだろうか。

コンピュータ処理能力の向上に伴い、大量データ（テキスト含むマルチモーダルデータ）の処理が可能になった一方で、そうした大量のデータの

処理にはまだ多くの課題がある。たとえば、単純な集計や統計処理などは機械的な処理になじむが、データから意味を見いだすというようなことはコンピュータにとっては現在でもまだ非常に困難な課題である。大量なデータの海から必要な情報を見つけ、見つけたデータをどのように可視化するか、といった知識活用の視点を情報システムに織り込むことは、企業の競争力を形成する非常に重要な要素になっていると言える。

また大量のデータの中に必ずしも求めるデータがあるとは限らないという課題も、実務的には頻繁に言及される。そこで、高質なデータを集めるための仕組み（ITプラットフォームのみならず、インセンティブをどうするか、などの制度的な問題も含む）を整備する重要性が増している。こうした課題に対し、情報システムを活用したナレッジマネジメントの構築が重要な役割を果たすようになってきている。

こうしたことから、情報システム利用の巧拙は、情報システムの構築にあたり、知識活用の視点を確実に埋め込めるかどうかにかかっているのではないかと考えることができる。こうした視点の具体像を明確にするため、以下、ナレッジマネジメントの先行研究についてレビューしていこう。

### 2-1 ナレッジマネジメントの先行研究

ナレッジマネジメントとは何か。この定義に関しては、2000年代前半をピークに、様々な論者により議論されてきたが、ここでは代表的なものとして、野中らの定義を紹介する。野中らによると、ナレッジマネジメントとは、「*個々人の知識や企業の知識資産 (Knowledge Asset) を組織的に集結・共有することで効率を高めたり価値を生み出すこと。そして、そのための仕組みづくりや技術の活用を行うこと*」と定義されている（野中1999）。ナレッジマネジメントは、知識を企業の現場で活かすための取り組み全般を指すものと言えよう。ここで、ナレッジマネジメントの推進を考える場合、知識とは何か、知識を活用する舞台はどこか、という問いが重要な意味を持つ。特に、本研究では情報と知識を弁別する。企業内で情報

の活用というときは、効率向上が主な目的であり、効果もその範囲内に留まるものである。これに対し、知識活用は、組織の持続的な価値創造を目的とし、経済的な付加価値の増大を視野にいれるものである。下記で、それぞれ具体的に見ていくことにしよう。

### 2-1-1 知識とは

本稿において知識とは何かについて定義することは非常に重要な作業である。知識の定義は大きく次の3つの文脈に沿って行うことができると考える。すなわち、哲学的な定義、認知心理学的な定義、知識工学的な定義である。

哲学的な定義に基づく研究のひとつとしては、野中らの知識創造理論などがある。野中らは、知識を定義するにあたり「正当化された真なる信念」というプラトン以来の西洋哲学の伝統的な定義を採用している（野中 1995）。野中らは、これまでの経済学や経営学の研究は、はっきりとは知識を捉えてこなかったか、あるいは「既存の知識」を扱うに留まっていたために不十分であると論じている。組織的知識創造理論は、知識を情報とはっきりと区別し、知識を「ダイナミックなプロセス」と捉える。また、「知識創造企業」となるには、知識担当役員（CKO）を置き、先端的情報技術で知識ベースを構築しているだけでは不十分で、製品開発のイノベーションだけでなく、日常的な問題解決、顧客との協調作業、などあらゆる局面で知識創造する企業であることが求められる、としている（伊丹 2000）。

認知心理学的な定義に基づく研究としては、松尾（松尾 2006）の研究などがある。心理学では、人間の記憶のうち、保持時間の長い長期記憶に貯えられる情報を知識とし、知識は宣言的知識と手続き的知識とに分かれるとする。宣言的知識とは、たとえば、「犬＝ペット」というようなある概念を定義づけるための知識である。また、手続き的知識とは、たとえば、犬の世話をするための知識のように、何らかの手続きを実行するための知識である。このうち、経営学においては主に手続き的知識に注目する。松尾は、営業マンが熟達する

プロセスにおける手続き的知識の重要性について論じている。

知識工学的な定義に基づく研究としては、溝口ら（溝口 1999）のオントロジー研究などがある。知識工学においては、知識を意志決定の推論過程に直接影響を与える情報と定義する。また、知識の定式化や知識の表現方法について考察を重ねる。ナレッジマネジメントが属する経営学では、こうした3つの文脈を織りなすようにして複合的に知識を捉えようとする傾向が強いようである。いずれにしても、この3つの文脈に共通しているのは、知識を扱うということは、単に情報を扱うこととは異なり、人間の知的な営みの程度がそれだけ高くなると考えている点である。本稿における知識の考え方も、基本的にこの3つの文脈を重層的に捉える立場であるが、もう少し具体的に知識の考え方について整理しておこう。

この整理に役立つのが図1である。図1の左側のピラミッドは、知識のとらえ方についての簡単な模式図を示したものである。知識は、使われる人や場所などといった文脈への依存性が高いため、一義的にその定義を定めたり具体例を示すことは難しいが、ここでは簡単に、知識という概念を、「データ」、「情報」、「知識」という3階層に分け、上位の階層にあがるに従って、意味の豊穡さが増すようなもの、と捉えてみる。

意味の豊穡さが増すにつれ、知識の扱い方や活用の仕方も難しくなると考えられるが、それぞれの階層における、知識の適切な扱い方を考えることは有用である。

知識活用の初期のステージは、主にデータを扱う段階であるといえよう。このステージにおいては、知識の活用レベルも低く、組織内に蓄積されている意味の豊穡さは低いと考えられる。このステージのデータは売上高など、主として数値データか、リレーショナルDBとして収集・蓄積されている。このデータが統計解析手法によって加工され利用される。

次のステージは、主に情報を扱う段階である。ここでは、データマイニングやテキストマイニングといった、データを加工／分析する技術が主役

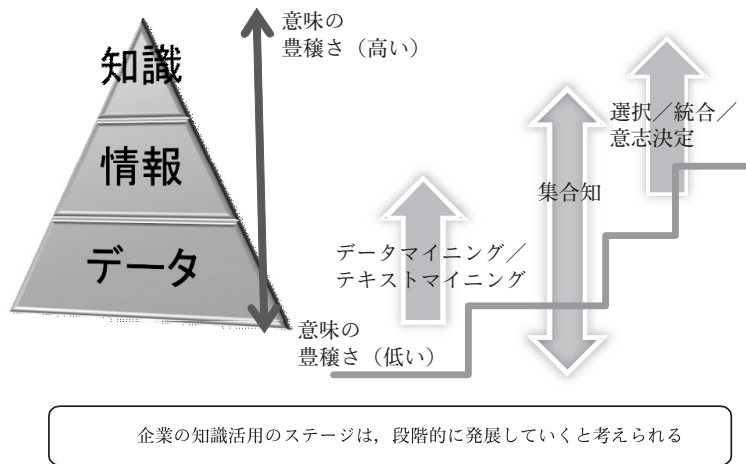


図1 企業の知識活用ステージ（模式図）

（出所） 筆者作成。

になる。組織内に蓄積される意味の豊穡さはまだ比較的低い段階にあるが、組織内のデータ／情報を効率的に扱うことができる状態になっている。次のステージは、知識を扱う段階である。知識の量や質が不足していると考えられるときには、集合知などの技術を用いて、知識の質や量を調整する。このステージにおいては、知識が選択／統合され、意志決定に活用されるなど、高度な活用が行われる。具体的には、たとえば、売上高などの業績データを、商品別、地域別、担当別、曜日別などのように分析のメッシュを変え、あるいは複合的に組み合わせることによって、これまでは知られていなかった地域固有の売上増大の法則が発見される、といったような活用である。結果として組織内に蓄積されている意味の豊穡さも高くなる。このように、企業の知識活用ステージは段階的に発展していくと考えられる。各ステージの特性に沿った、知識活用戦略を用いることが重要であるといえよう。

本稿の目的はテキストデータをコンピュータと統計解析手法を用いて処理し、意味を発見するための改善手法を提示することにある。したがって、ここで、知識を扱う上でのコンピュータ利用の可能性について整理しておこう。

ティアニーらは、知識の活用戦略の類型として、「コード化戦略」と「個人化戦略」という概念を

提案している（ティアニー 1999）。ティアニーらによると、「コード化戦略」は、様々な知を慎重にコード化し、データベースに入力し、従業員が容易にアクセスできるようにすることであり、知識は再利用できるという考え方に基づくものと言える。また、「個人化戦略」は、知とそれを生み出した人物を不可分と捉え、コンピュータの役割は、知を交換しあうための補助的なものと捉えることとされ、専門性を志向するものと言える。ティアニーらは、上記の戦略のどちらを主軸に据えるかは二者択一（どちらも同じように採用することは望ましくない）であると考えており、80：20の比率でどちらかにリソース配分することが、経験上もっとも効果的であると述べている。知識を扱う上でコンピュータを利用することは、コード化戦略を推進する場合に効果的であると言えるが、知の交換がサイバー空間で行われていることが普通である現代の企業環境においては、個人化戦略をとったとしてもコンピュータを効果的に使うことは可能であろう。

本稿においては、知識を単なる情報とは峻別した、重層的な意味の複合体と考えるが、知識を扱う上でコンピュータ利用を不可欠のものとする立場から、知識の一部はコード化可能なものである、と考える。

### コード化戦略

様々な知を、慎重にコード化し、データベースに入力し、従業員が容易にアクセスできるようにすること。

### 個人化戦略

知とそれを生み出した人物を不可分と捉え、コンピュータの役割は、知を交換しあうための補助的なものと捉えること。

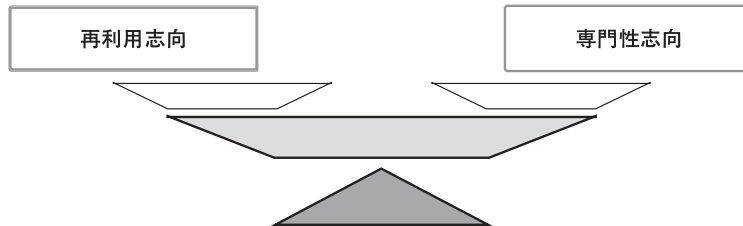


図2 コード化戦略と個人化戦略のバランス

(出所) ティアニー 1999 より筆者作成。

## 2-1-2 場とは

知識創造や情報共有が起こる舞台である「場」については、「組織」とは異なる概念として様々な識者がその重要性を論じている。

遠山ら（遠山 2000）は、「場」を、物理的な場所という意味以外に、空間と時間を同時に含む場所性の概念と捉えており、「関係」の空間という意味を内包しているものとしている。また、「場」が知識創造プロセスにエネルギーを与え、生み出される知識の質を決定する、と述べている。伊丹らは、こうした考え方を次のように端的に示している。

「創造する力」は、単に個人の内にあるのではなく、個人と個人の「関係」、個人と環境の「関係」、すなわち「場」から生まれる」（伊丹 2000）

また、伊丹は、「場」を次のように定義し、「場」によって、(1)人々の間の共通理解（組織の中の協同に非常に必要）が増すこと、(2)人々がそれぞれに個人としての情報蓄積を深めること、(3)人々の間の心理的共振が起きること、という効果が生み出されると述べている（伊丹 2005）。

「場とは、人々がそこに参加し、意識・無意識のうちに相互に観察し、コミュニケーションを行い、相互に理解し、相互に働きかけ合い、相互に心理的刺激をする、その状況の枠組みのことである」

このように、多くの経営学研究では、「場」を

物理的なものだけとは捉えず、「時空間」や「関係」を含んだ、複合的な概念として捉えている。

「なぜ企業はうまく情報が共有できないのか」、「情報共有が一時的に成功しても、継続しないことが多い」という素朴な実務的な課題については、「場」がうまく機能していない、もしくは、しなくなったという原因が関係していることが多い。これが経営学研究において「場」の重要性が議論されるゆえんである。

こうした議論の延長線上では、多くの論者がインターネットに代表されるサイバー空間をも「場」の射程内に捉えている（野中らの「システム場」（野中 1999））。サイバー空間は、グローバル化による組織の物理的な広がりや、働き方の多様性による従業員同士の時間的な接点の減少という極めて現代的な課題を念頭においたとき、不可欠な環境であると言えよう。物理的な「場」は知識創出にとって非常に重要ではあるが、現代においてはリアルな「場」というとらえ方だけでは不十分であると言えよう。

## 2-2 ナレッジマネジメントの新しい展開

### 2-2-1 Enterprise 2.0

以上見てきたように、ナレッジマネジメントは、企業実務における IT の利用を、非常に重要なものと捉えてきた。近年では、さらに、インターネット技術を使った、ユーザー参加型の情報システム



活用が目されている。「Enterprise 2.0」という考え方である。Enterprise 2.0は、本概念の提唱者である McAfee 教授の定義によると以下のようなものである。

「Enterprise 2.0 is the use of emergent social software platforms within companies, or between companies and their partners or customers.」

(筆者訳:「Enterprise 2.0 は、企業内、企業間、あるいはパートナーや顧客たちとの間での、新興のソーシャルソフトウェアプラットフォームの利用である」)

換言すると、インターネット技術を利用したサイバー空間の中で、「関係性」を導入したソフトウェアを利用することがその特徴とされている。「関係性」を導入したソフトウェアとは、たとえば、ある個人のページには知り合い情報、趣味の情報、スキルなどが登録されていて、その個人と他の個人の関係性はリンク構造でつながっており、組織内の人のつながりをクリックでたどっていきけるようにするようなソフトウェアである。そして、McAfee (McAfee 2006) によれば、Enterprise

2.0には5つの特徴がある。表1にはその内容を示す。

近年インターネット上で爆発的な普及を遂げている SNS (Social Network Service) サイトの思想が、社内のイントラネットの世界でも有効、ということであろう。Enterprise 2.0に関する取り組みは近年多く発表されており、代表的な取り組みについて表2にまとめた。厳密に言えば、5つの特徴の全てを備えているわけではない取り組みもあるが、広義の意味におけるソーシャルソフトウェアの導入が効果を上げている事例として、どれも興味深い。

上段3つめまでの事例は、主に社内業務プロセスの効率化や情報共有に関するものであり、日本ではこのような使われ方が比較的多い。他方、下段2つの事例は、顧客まで取り込んだ知識創発の事例であり、米国などではこのような形の事例が比較的取りざたされることが多い。ここに挙げられた事例は成功例ではあるが、社内の情報共有や知識創造に対し、Enterprise 2.0が有効であることは徐々に認知されているところであろう。

マカフィーが指摘するように、「エンタープライズ2.0と伝統的なアプローチは二者択一ではなく、強化し合う関係にある」(マカフィー2010)。

表1 Enterprise 2.0の特徴

特 徴	内 容
S (Search)	必要な情報の検索。専門家によって情報が構造化されていれば、検索は容易になる。
L (Link)	情報の関係付け。規模の限られたイントラネット空間では、人々が自らリンクを貼れるようにすべきである。
A (Authoring)	文章の作成、構成。Blogのように人々が自ら情報を発信、公開できるようにすることが創造性と継続性を生む。
T (Tag)	情報のカテゴリズ。適切なタグの作成により、従業員同士のナレッジワークの様子が可視化され、有用な情報が利用可能な状態でつなぎ止められる。
E (Extension)	コンピュータが自動でカテゴリズやパターンマッチングを行うことで、ユーザーにとってのタグ付けの価値を一步先まで高めること。
S (Signal)	ユーザーが興味のある情報が更新されたかどうかをわざわざチェックするのではなくシステムが自動で通知してくれること。

表2 Enterprise 2.0 の取り組み事例

事 例	事 例 内 容
東芝：メッセージ集約 コミュニティウェア GroupScribe	コミュニケーション情報の活用ツールを全社で運用し、プロジェクト単位での情報共有や週報の共有で効果を上げている。
NTT データ：Nexti (社内 SNS)	2006 年 4 月に正式に運用が開始され、2009 年 2 月時点で約 7,400 人が登録。Q & A コーナーでは、平均して 1 日 1 件以上の質問が投稿され、回答が複数寄せられる。また、コミュニティ数は約 1,000 (半数強は業務に関連したカテゴリー)、日記は 1 日に約 100 件のエントリーがある。
IBM：イノベーション ジャム	IBM グループの「全世界全社規模のネット会議」。2006 年 7 月には、72 時間にわたる 2 回のセッションで、同社の経営革新を図るための 4 万 6,000 件以上のアイデアが提出された。11 月にはこれらのアイデアから 10 種類の新ビジネスが厳選された (村上・武田 2007)。
P & G：コネクト・ア ンド・ディベロップ	オープンイノベーションを促すサイトを立ち上げた結果、R & D 効率が約 60% 向上し、対売上高研究開発費率も 30% 近く低下した。そして、全社のイノベーションの 35% 超、売上高にして数十億ドルを生み出している (ヒューストン 2006)。
DELL：Idea Storm DELL	DELL の顧客を対象にしたウェブサイト。2 年間で 9,000 件以上のアイデアに対して 60 万件以上の投票がされ、200 件以上の製品改善が実現された。

そのため、Enterprise 2.0 の取り組みがあるだけでは不十分で、従来行われているように、知識データベースを作成、メンテナンスし、従業員の知識再利用を促すことは、依然として重要な要素であろう。ただ、Enterprise 2.0 の考え方は、組織における知識の質を維持するために非常に重要である。情報が企業の中に散在する現代では、その情報を集約し、体系化し、再利用可能な形で整理しておくこと (=このような取り組みを「知識化」と言い換えてもよいだろう) が、知識の活用にとって最も重要な要素であると考えられるからである。

筆者が提唱することは、場の限界を補完するためには、コンピュータネットワーク上での知識創出、とりわけテキストマイニングによる知識創出が役立つということである。ただ、テキストマイニングそれ自体には限界があるので、テキストマイニングに Enterprise 2.0 の要素を取り入れるということである。

## 2-2-2 テキストマイニング

上述したように、IT を活用して知識の質を維持する試みの進展が、最近のナレッジマネジメントの新しい側面の一つである。さらに、それをどのように活用するかということについての新しい側面として、近年、テキストマイニングが注目されている。テキストマイニングとは、数値データからのみではなく、大量のテキストデータから人間に有用な知識を発見する手法の総称である。実務界では、コールセンタのテキスト分析によるクレーム兆候の検知や、アンケートデータの解析によるマーケティング高度化などで多くの事例が報告されている。以下、テキストマイニングの先行研究について代表的なものを概観しておこう。

(A) 松村ら (松村 2007) は、顧客から寄せられたメッセージには表出していない背後のニーズを明らかにするために、テキストマイニングを用いて分析を行っている。本先行研究における新規性は、様々な顧客がメッセージのやりとりをして

いる掲示板のようなサイトにおいて、これまで語られていなかった新たな情報が語られ出す契機となったメッセージを抽出する特徴度を定義して、その有用性を示したことである。その際、文の意味構造を、恣意性を排除して把握するための辞書として、用言を用いた格フレームを使用していることが特徴である。

(B) 喜田(喜田2007)の研究では、有価証券報告書に出現する名詞概念と経営成果(シェア, 売上高, 経常利益)や一般経済環境, 業界環境の間の相関分析を行っている。結果として、アサヒビールについては経常利益, 売上高, 業界環境とは正の相関があり, 一般経済環境とは負の相関があること, キリンビールについては経常利益, 売上高, 業界環境と正の相関があり, シェア, 一般経済環境とは負の相関があることが示されている。本先行研究が、テキストマイニングにおいて名詞に着目して分析を行っていることは、認知心理学の分野では、名詞の数の変化および文字数の増加は学習(認知変化)であるとされていることに根拠をおいている。

(C) 三菱総研の小林らの研究(小林2004)では、インターネットリソース(特に、IT系のニュースメディアであるITmediaのニュースコンテンツ)を使い、情報通信分野の技術マップを自動生成する手法を提案している。

(D) 那須川(那須川2006)は、PCメーカーのヘルプデスク業務にテキストマイニング技術を取り入れ、業務効率改善に成功したという事例を報告している。

(E) 西山ら(西山2009)は「何ができる技術か?」ということに着目して技術の有効性を判断するための手法を提案している。具体的には、(1)「当該技術の新たな長所を示した」特長表現を、「増強クラス」と「改善クラス」に大別し、それぞれの特徴的な語尾に着目して、名詞を抽出すること、および、(2)その名詞を当該技術分野では珍しく、かつ一般的な名詞であった場合に高い順位をつけるというランキング方法をとること、の2点を特徴とするもので、特許データと製品発表データをもとにした実験でその有用性を確認している。

(F) 市村ら(市村2001)は、ライオン株式会社において販売プロモーション活動を行う「フィールドレディー」が日々書き留める日報を対象にテキスト分析を行い、チャンスロスを低減するための仕組み作りを行った。日報には、製品の売れ行き、陳列状況、店員の評判や対応などが記されており、日報管理者は、これらの日報すべてに細かく目を通し、重要な部分を手作業で抽出して加工し社内の文書をまとめていたが、本先行研究の行った施策によって日報管理者の負荷が大幅に低減されたようである。本先行研究の施策は、大量の日報を、自社情報、競合他社情報、エリア別情報という分類軸に基づいて分類し、自動的にレポートングをするというもので、あれば売れていたのに欠品していたために機会を失ったということを低減する、効率的な情報取得の仕組みを構築した事例として興味深いものである。

### 2-2-3 テキストマイニングの方法論としての強み

以上見てきたように、テキストマイニングは、経営課題の解決に対して実践知識を提供することにより、一定の貢献をするものと捉えることができるだろう。図3に示すように、テキストマイニングの方法論としての強みは、大量の情報の海から効率よく効果的に意味を把握できること、にある。単なる数値の集計だけではわからない意味の把握を、地道なインタビューの積み重ねのような高いコストを要する方法だけに頼らず、効率的に行うことができることこそテキストマイニングの強みなのである。

しかし、先行研究の限界は、テキストマイニングを全ての経営プロセスにおいて、全ての従業員が容易に扱えるほど、企業実務にテキストマイニングを埋め込めていない点である。(A)から(C)の研究は、テキストマイニングによって組織内で起こる様々な現象の分析を行うものであるが、それによって即、日常の経営課題の解決につながるものではない。また、(D)から(F)の研究は、企業実務の課題解決を狙っているものではあるが、テキストマイニングの運用は、あくまでテキストマイニン



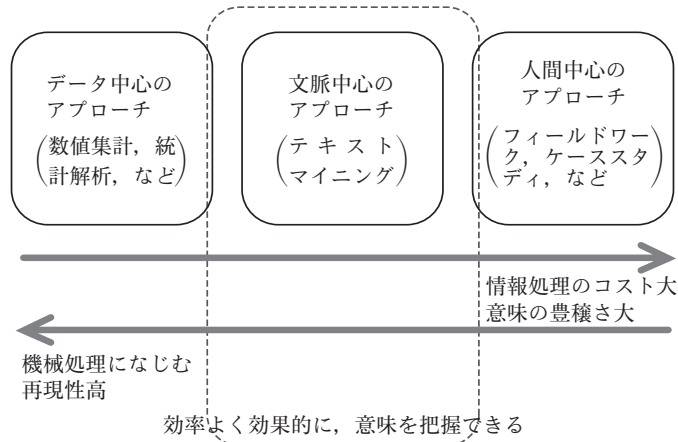


図3 テキストマイニングの方法論としての強み（概念図）

（出所）筆者作成。

グ研究者が一時的に行っているものであって、普通の従業員が普通にテキストマイニングを使って課題を解決するプロセスを洗練できているわけではない。企業の実際の現場では、普通の従業員が、テキストマイニングなどの知識を分析するツールを普通に用いて、しかもそれを継続させなければならない。

第3項では、こうした限界を克服するために、Enterprise 2.0 の考え方を取り込んだテキストマイニングの可能性について、述べていく。

### 3. 仮説検証（Enterprise 2.0 の考え方を取り込んだテキストマイニングの可能性）

#### 3-1 テキストマイニング手法洗練のための諸条件

これまでの先行研究のレビューを踏まえて、テキストマイニング手法の洗練に向けた知識活用手法の構築に関する諸条件を考察していこう。

第一に、企業内に散在する情報が、知識化されていることが必要である。これは伝統的なナレッジマネジメント研究からの知見であるが、知識を確実に形式化し、しかもそれが専門家によって不断にメンテナンスされていることは、多くの研究においてナレッジマネジメント成功の要素とされている。多くの場合、その情報の一番の専門家は、

オリジナルデータの生成者であり、業務担当者である。第二に、それらの知識が体系化されていることが必要である。インターネットでの各種SNSサービスの興隆に見るように、Enterprise 2.0 の思想に基づいて知識を体系化することは企業内においても重要なナレッジマネジメントの成功要素である。第三に、体系化された知識が、不断に分析され、再知識化、再体系化されていることが必要である。継続的なイノベーションのためには、このような継続を促す仕組みが必要である。

#### 3-2 実証

本項では、前項で述べた諸条件の妥当性について、簡易な検証を試みる。

##### 3-2-1 実証方法

JST で構築を進めている、失敗知識データベースの一事例を、テキストマイニングで分析する。失敗知識データベースは、上記諸条件のうち、第一、第二、第三の全てを満たしていると言えるが、このうち、第一の部分について、統制を加え、上記提唱した諸条件第一を満たす場合と、満たさない場合の2つのテキストを準備し、それぞれの分析結果にどのような違いが生じるのかを観察する。

失敗知識データベースを、企業の知識データベースのサンプルとして選定した理由については、次

の通りである。(1)失敗知識データベースは、平成17年の一般公開以来、継続的に専門家によってメンテナンスされている質のよい知識データベースであること、(2)失敗知識データベースは、平成21年度には、約513万アクセス（日本語版）を数え、一般的にも有用な知識データベースと認知されていること、(3)失敗知識データベースは、検索容易性、情報タグ付けに優れた知識データベースであり、Enterprise 2.0の考え方を一部取り入れており、本目的の趣旨に添った分析を行うために適切であること、の3点である。

### 3-2-2 マイニング対象

失敗知識データベースの「みずほフィナンシャルグループ大規模システム障害」事例を対象とする。

なお、図4は、本分析の対象となる失敗知識データベースの当該事例のページの一部である。

### 3-2-3 マイニング方法

フリーソフト「KH Coder」を用い、共起（同一文章での同時発生）情報を元にしたネットワークグラフ生成により、マイニング対象に含まれる「知識」の構造を明らかにする。

具体的には、上記で提唱した諸条件を満たすよ

うなテキスト集合と、満たさないようなテキスト集合を準備し、それぞれのテキスト集合に対してテキストマイニング分析を行う。上記で提唱した諸条件を満たさないようなテキスト集合とは、事例の文章項目全体のうち、「事実に関する項目」だけを抜き出した文章群（下記(b)の項目群）であり、満たすようなテキスト集合とは、「事実+知見に関する項目」を抜き出した文章群（下記(a)の項目群）を指す。後者のテキスト集合は、いわば知識化済み事例であり、前者のテキスト集合はいわば事例情報に過ぎないといえることができる。これら2つのテキスト集合を分析した結果として、知識表現の構造がどのように違うかを明らかにする。

(a) 上記で提唱した諸条件を満たす場合の対象項目

事例名称、事例発生日付、事例発生地、事例発生場所、事例概要、事象、経過、原因、対処、対策、**知識化**、**背景**、**よもやま話**、情報源、備考、分野、データ作成者

(b) 上記で提唱した諸条件を満たさない場合の対象項目

事例名称、事例発生日付、事例発生地、事例発生場所、事例概要、事象、経過、原因、対処、対策、情報源、備考、分野、データ作成者

失敗事例

戻る


シナリオ	目録	失敗百選 (PDF)
事例名称	みずほフィナンシャルグループ大規模システム障害	
代表図		
事例発生日付	2002年04月01日	
事例発生地	全国	
事例発生場所	みずほ銀行	
事例概要	第一勧業、富士、日本興業の3銀行のシステムを「みずほ銀行」として一本化するシステム統合で、統合の方針決定が好余曲折し、システム統合のスケジュール・統合作業が遅れて、予定していたシステム稼働テストの開始がずれ込み、十分な見極めができないまま開業したため、開業初日から現金自動預入払出機(ATM)の障害が発生。さらに公共料金の自動引き落としなどの口座振替に遅延が生じるトラブルが起きた。トラブル発生後も対応が遅れるなどで、口座振替の遅延が拡大、大混乱となり最大級のシステム障害に陥った。	
事象	第一勧業、富士、日本興業の3銀行が再編して誕生した「みずほ銀行」は、開業初日(2002年4月1日)に現金自動預入払出機(ATM)の障害が発生。さらに公共料金の自動引き落としなどの口座振替に遅延が生じるトラブルが起き拡大した。口座振替に遅延は14月1日だけで105,000件に達し、翌日以降の積み増しとなり、連続的に大量の未処理が発生。4月6日(口座振替開始)250万件にまで増殖していった。それに付随したかたわで、約3万件の二重引き落とし発覚し最大級のシステム障害に陥った。旧慣を固るも、このトラブルで顧客の混乱が長期化、1ヶ月以上も続いた。	

図4 JST失敗知識データベースの事例紹介画面

(出所) JST失敗知識データベース (<http://shippai.jst.go.jp/fkd/Search>)

### 3-3 結果

図5は、上記(a)と(b)の条件によるテキスト集合をテキストマイニング分析した結果を比較したものである。

(b)の事例のみのテキスト集合においては、「リレーコンピュータ」「システム稼働テスト」「負荷テスト」などの語句が右中段に現れており、今回の失敗事例における、システム構築面での失敗原因に関する理解を助けることができるものと期待できる。しかしながら、これだけでは、「どのように構築したために失敗したのか」といった人間的側面についての理解が不十分である。

(a)の知識化済みのテキスト集合をテキストマイニングした例では、「たすき掛け」「主導権争い」「意思決定」「経営統合」といった語句群が中央下部に現れており、システム構築における人間的側面での失敗原因を理解する上で役に立つ知識表現になっていることがわかる。(a)のテキスト集合には、専門家がメンテナンスした項目である「知識化、背景、よもやま話」といったテキストが含まれており、こうした Enterprise 2.0 的な考えに基づく知識化済みテキストを整備することが、テキストマイニング分析の新しい展開に向けたひとつ

の重要な要素であるといえるだろう。

### 4. まとめと課題

本稿では、企業実務における実践的な知識の活用方法について、Enterprise 2.0 の考え方を取り入れたテキストマイニング分析の新たな展開について考察を加えてきた。伝統的なナレッジマネジメントの手法は否定されるべきものではないが、インターネットで進化の著しい SNS などの技術や思想を、社内イントラネットで展開することは価値のある取り組みであるように思われる。

また、従来のテキストマイニング分析のように、文章集合を所与のものとすると、分析結果の広がりには限界があるが、Enterprise 2.0 の考え方を取り入れて整備された知識データベースのテキスト集合をテキストマイニングすることは、分析結果に広がりをもたらすことがわかった。具体的には、3-1 で述べた諸条件を満たす場合に、テキストマイニング分析の結果が、より人間的側面を捉えた、洗練されたものとなるという仮説の妥当性について一部検証できたものとする。本稿で述べた「知識化」という概念は、知識工学研究などというアノテーション付与とほぼ同義であると言える

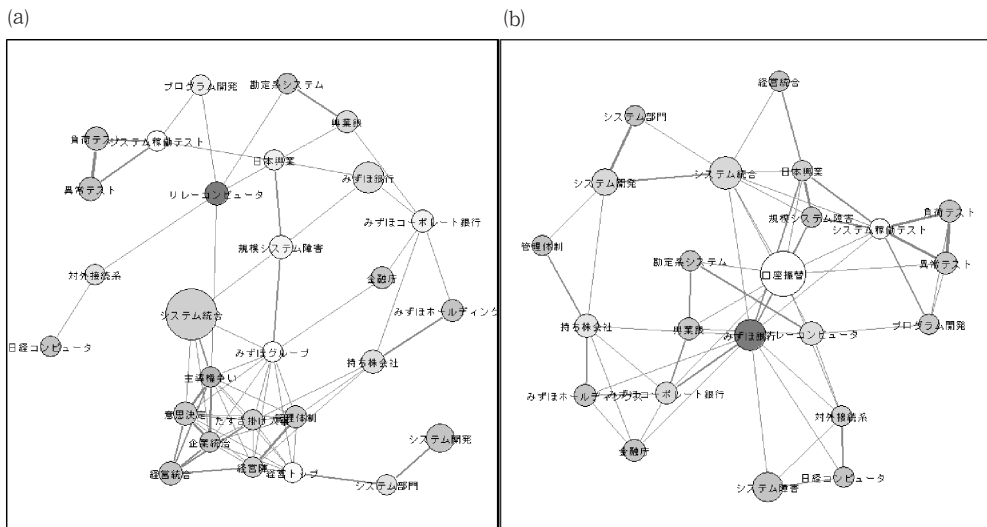


図5 テキストマイニング結果

(出所) 筆者作成。

が、アノテーション付与に関連するインターネットサービスの豊富さは、この考え方の方向性が間違っていないことを示す傍証とできるであろう。

ここで、3-3で得られた検証結果を本稿の目的との関連で解釈してみよう。本検証で情報源として利用した失敗知識データベースは、いわば日常業務の継続的な改善活動の見本である。本検証結果からは、企業の継続的なイノベーションを支えるためには、失敗知識データベースのような日常業務における知識データベースを整備すること、およびそこから知見を引き出すためのテキストマイニングの活用が重要であるということが実証できたと言える。

しかしながら、本稿ではまだ検証しきれなかった不十分な点がある。一つは、テキストマイニング適用の前提となる、文章解析のための辞書整備について触れなかったことである。これについてはテキストマイニングに関する諸研究において詳しいため割愛したが、ナレッジマネジメントを新たな局面に導くために、どのような辞書整備を行うべきかという課題を追求することは非常に重要な作業であり、今後の課題と考えている。またもう一つには、テキストマイニング手法自体をブラックボックスとして扱い、テキストマイニングのプロセス自体に言及していない点である。多くの企業実務者にとって、テキストマイニングのアルゴリズムを理解しながら当該手法を使うということは現実的ではないとも思われるが、本稿で述べた視点を取り込みつつ、よりよい結果を導くための最適なアルゴリズムについて考察することは重要であり、今後の課題としたい。最後に、本研究を実際の企業実務の現場に適用して効果を検証することは大きな課題であると考えている。筆者の提唱する「知識化」は、単なるアノテーション付与を超えて、人間と人間、人間と機械、機械と機械、といった様々なインターフェースにおける対話を、より効果的に行うことを目指しており、今後、コンピュータネットワーク上で、こうした考え方を、実際のフィールドで適用することが大変重要な知見をもたらすことになるように思われる。こうした展望を踏まえ、今後、これら3つの課題に取り

組んでいく。

#### 参考文献

- McAfee, A. P. (2006), "Enterprise 2.0: The Dawn of Emergent Collaboration," *The MIT Sloan Management Review*, Vol. 47, No. 3, 2006.
- ティアニー, T.・ノーリア, N.・ハンセン, M. T. (1999), 「コンサルティング・ファームに学ぶ『知』の活用戦略」『DIAMOND ハーバード・ビジネス・レビュー』, 1999年9月号, pp. 60-74, ダイアモンド社.
- マカフィー, A. P. (2010), 「エンタープライズ 2.0 がイノベーションを変える」『DIAMOND ハーバード・ビジネス・レビュー』, 2010年4月号, pp. 81-83, ダイアモンド社.
- 伊丹敬之・野中郁次郎・西口敏宏 (2000), 『場のダイナミズムと企業』, 東洋経済新報社.
- 伊丹敬之 (2005), 『場の論理とマネジメント』, 東洋経済新報社.
- 市村由美・鈴木優 (2001), 「テキストマイニング技術と応用」『東芝レビュー』, Vol. 56, No. 5, pp. 19-22.
- 小林慎一・白井康之・糸野文洋・犬島浩 (2004), 「インターネットリソースを活用した情報技術マップ構築支援環境」, 2004年度人工知能学会全国大会(第18回)論文集.
- 遠山亮子・野中郁次郎 (2000), 「『よい場』と革新的リーダーシップ：組織的知識創造についての試論」『一橋ビジネスレビュー』 Vol. 48, Sum.-Aut., pp. 1-13.
- 那須川哲哉 (2006), 『テキストマイニングを使う技術／作る技術——基礎技術と適用事例から導く本質と活用法』, 東京電機大学出版局.
- 西山莉紗・竹内広宣・渡辺日出雄・那須川哲哉 (2009), 「新技術が持つ特長に注目した技術調査支援ツール」『人工知能学会論文誌』 Vol. 24, No. 6, pp. 541-548.
- 野中郁次郎・竹内弘高 (1995), 『知識創造企業』, 東洋経済新報社.
- 野中郁次郎・紺野登 (1999), 『知識経営のすすめ』, ちくま新書.
- 松尾睦 (2006), 『経験からの学習：プロフェッショナルへの成長プロセス』, 同文館出版.
- 松村真宏 (2007), 「メッセージの背後に潜む「問い」の抽出」, 『人工知能学会論文誌』 Vol. 22, No. 1, pp. 93-102.
- 溝口理一郎 (1999), 「オントロジー研究の基礎と応用」『人工知能学会誌』 Vol. 14, No. 6, pp. 977-988.

村上明子・武田浩一 (2007), 「企業における集合知の活用事例「Innovation Jam」」『ProVision』, No. 55, pp. 26-32.

喜田昌樹 (2007), 『組織革新の認知的研究 — 認知変化・知識の可視化と組織科学へのテキストマイニ

ングの導入』, 白桃書房.

ヒューストン, L.・サッカブ, N. (2006), 「P & G : コネクト・アンド・ディベロップ戦略」『DIAMOND ハーバード・ビジネス・レビュー』, 2006年8月号, pp. 45-56, ダイアモンド社.