

ナレッジマネジメントにおけるテキストマイニング

松井くにお

(株) 富士通研究所
matsui.kunio@jp.fujitsu.com

内野 寛治

(株) 富士通研究所
kanji@jp.fujitsu.com

渡部 勇

(株) 富士通研究所
watanabe.isamu@jp.fujitsu.com

SECI モデルを回すには

ナレッジマネジメントの連載は、第5回目を迎えますが、いろいろな業務に適用するにあたって、SECIモデルを回すことが共通した知識創造の原則でした。SECIモデルを暗黙知と形式知に焦点を当てて書き直すと、表出化・内面化などの「知」の変換は、プロセスとして描くことができます(図-1)。現実を見てみると、暗黙知・形式知は、それぞれのレベルで「匠の技伝承」や「事例の再利用」という形で活用されています。また、事例の蓄積、情報の組織化など、暗黙知から形式知への表出化のプロセスは盛んです。しかし、大量の形式知を持て余すあまり、形式知から暗黙知への内面化のプロセスが十分ではありません。これがSECIモデルを回す障害になっているのではないかと思います。

では、この内面化を活性化するためにはどうしたらよいのでしょうか。散在する大量の形式知には、重要な情報が埋もれて見えなくなっています。そこで、内面化のプロセスの途

中に「分析知」という知識を位置付けることにより、内面化支援の可能性を、ITと対話を用いて試みてみました(図-2)。この「分析知」を産み出すトリガーは、対話やコンサルテーションであり、分析のための視点を与えることができます。さらに大量の形式知を統合・分析するためには、テキストマイニングの技術を用いて「見える化」を行うことで、暗黙知を刺激する「気づき」を与えることができます。

□テキストマイニングとその応用

ナレッジマネジメントを支える基盤技術の1つとして「テキストマイニング」があります。テキストマイニングとは、文書情報から有益な知識を発見・抽出するための技術であり、数値・カテゴリ情報を対象としたデータマイニングに自然言語処理・情報検索技術などを統合する形で発展を遂げてきています。情報検索システムが、利用者の目的に合った文書を探し出すことを目的としているのに対して、テキストマイニングでは、文書を個別に調べても分からない、文書群全

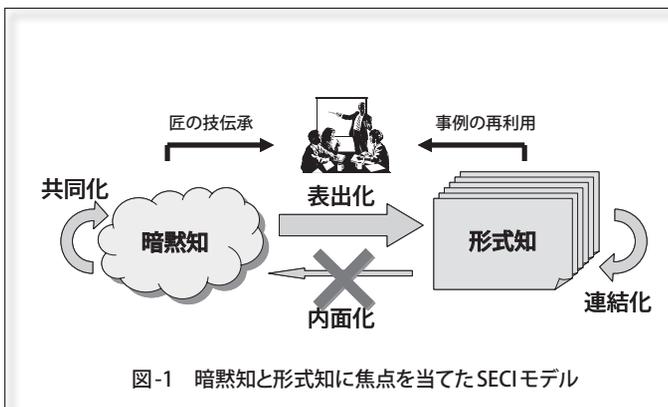


図-1 暗黙知と形式知に焦点を当てたSECIモデル

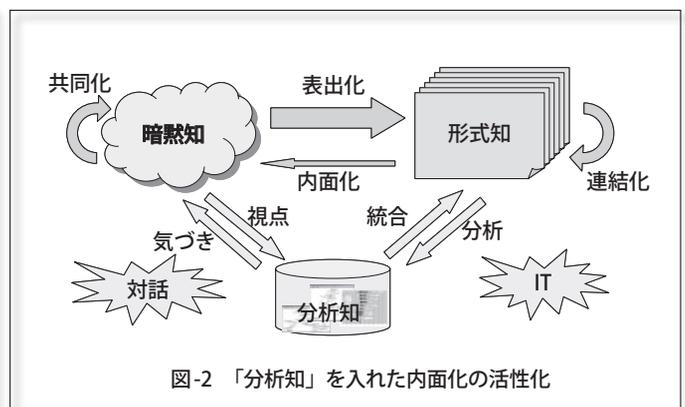
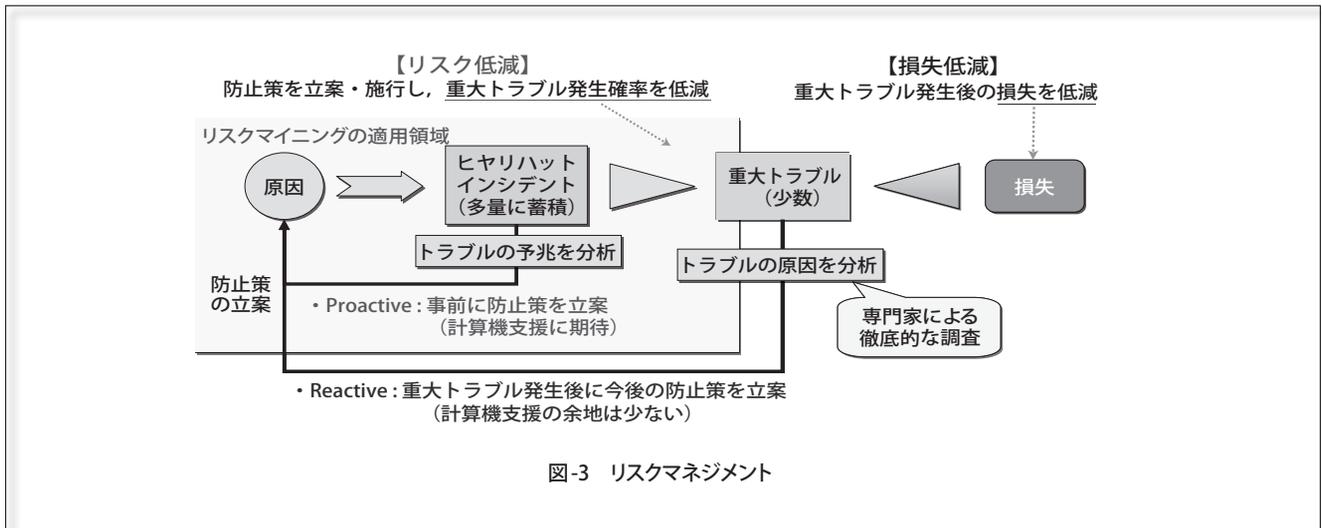


図-2 「分析知」を入れた内面化の活性化



体に内在する知識（パターンやトレンド）を発見することを目的としています。まだ新しい研究領域ですが、この10年ほどの間に実用化が進み、大量のテキスト情報にアクセスするための新しい道具として、ビジネスの場面でも活用されるようになってきています¹⁾。

富士通研究所では、自然言語処理技術と可視化技術を用いることによりユーザの知識発見を支援するビジュアルテキストマイニング技術^{2), 3)}を開発し、フリーアンサーのアンケート分析、顧客からの問合せ情報の分析、特許データの分析や評価⁴⁾などに適用する研究を行ってきました。近年では、リスクマネジメントにテキストマイニング技術を適用する「リスクマイニング」の研究を進めており⁵⁾、日本航空、東京大学と共同で航空分野への適用を行っています⁶⁾。また、ブログや掲示板、SNSなどのCGM（Consumer Generated Media）を分析し、「Webマーケティング」に応用する研究も進めています。本稿では、こうした「分析知」を用いた事例として、「リスクマイニングへの応用」と「Webマーケティングへの応用」について紹介します。

リスクマイニングへの応用

近年、企業経営における「リスクマネジメント」の重要性がクローズアップされてきています。ひとたび重大トラブル（事故や障害）を引き起こすと、金銭的な損失はもちろんのこと、企業の信用やブランドの失墜により、企業の存亡の危機に陥ることも少なくありません。「安全・安心」は、医療、鉄道、航空、原子力といった特定分野にとどまらない全業種・業務に共通するキーワードとなってきています。

リスクマネジメントは、図-3に示すように、トラブル発生後の対処により損失を低減する「損失低減」と、トラブル防止策を立案・施行してトラブル発生確率を低減する「リスク低減」の2つの側面を持っています。リスク低減は、さらに、事後対応的に防止策を立案・施行して重大トラブルの

再発を防止する「Reactive」なアプローチと、事前対応的に防止策を立案・施行して重大トラブルを未然に防止する「Proactive」なアプローチに分けることができます。

重大トラブルが発生したときに、事故調査委員会などを組織して徹底的な原因調査を実施するというのが典型的なReactiveなアプローチです。Reactiveなアプローチにおいて重要なのは、調査を実施する専門家の知識やスキルであり、次に説明するProactiveなアプローチに比べると、計算機支援の余地はあまりありません。重大トラブルは頻繁に発生するわけではないので、統計的な手法がベースとなっているマイニング技術の適用には向いていないのです。

一方のProactiveなアプローチでは、重大トラブルになる手前のインシデント（軽微なトラブル）やヒヤリハット（幸運にもトラブルには至らなかったケース）の事例を収集・分析することにより、重大トラブルに結びつく可能性のある要因やパターンを見つけ、重大トラブルの発生を未然に防止することを目指します。ハインリッヒの法則（リスクマネジメントの分野における有名な経験則）によれば、1件の重大トラブルの背後には、29件のインシデントと300件のヒヤリハットが存在すると言われており、収集・分析に利用可能な事例の件数は多く、計算機支援への期待も大きなものとなっています。リスクマイニングでは、この収集・分析の過程をテキストマイニング技術によって支援することにより、客観的かつ定量的なリスクマネジメントを効率的に実現することを目的としています。

リスクマイニングのプロセス

図-4に、リスクマイニングのプロセスを示します。

データ収集

インシデントやヒヤリハットといったネガティブな事例データを収集するのは容易ではありません。特に近年では、ハー

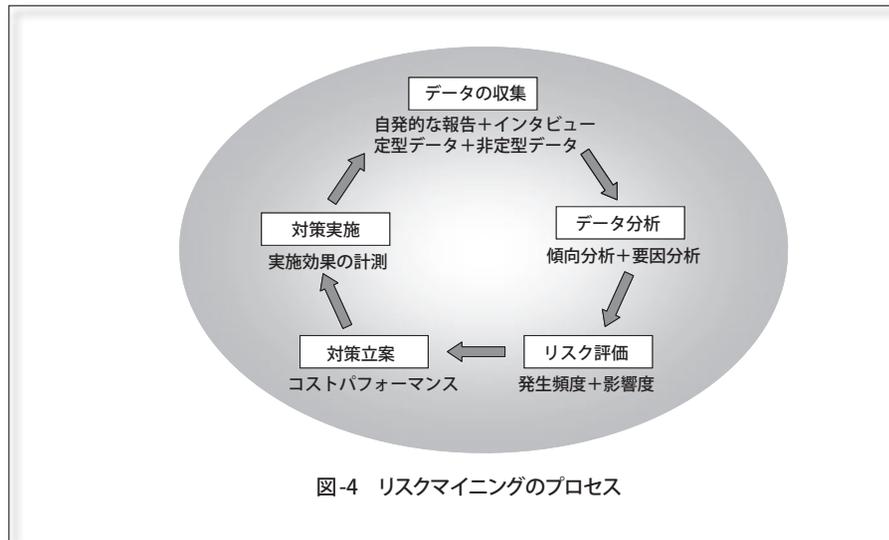


図-4 リスクマイニングのプロセス

ド面、ソフト面での安全対策が進み、トラブル原因に占めるヒューマンエラーの割合が増大してきているため、自らの失敗を報告することに対する心理的な障壁を取り払うための組織的な取り組みが不可欠となります。また、自発的な報告だけに頼っていたのでは客観性の点で限界があるため、行動・状況をモニタリングすることによってインシデントやヒヤリハットの発生を検出し、当事者にインタビューするといったデータ収集の方法も検討する必要があります。

収集される事例データは、通常、数値情報やカテゴリ情報などの定型データと、テキストで記述された非定型データから構成される半構造的なデータとなります。トラブル分析に不可欠となる発生条件、現象、原因、対策といった詳細情報はテキスト部分に記述されているため、自然言語処理技術やパターン処理技術などを用いて情報抽出を行う必要があります。

◆データ分析

データ分析では、まず「傾向分析」を実施して、マクロな視点からインシデント発生の全体傾向やトレンドを俯瞰し、着目すべきインシデントを絞り込みます。次に「要因分析」を実施して、ミクロな視点から因果関係の連鎖・関係性を調べ、元データ（レポート類）も参照しながらインシデントの発生パターンや仕組みを解析していきます。

図-5, 6, 7は、航空安全レポートを分析した事例です。

図-5は、航空トラブルの現象と原因の間の因果関係をネットワーク図として可視化したもので、「BIRD STRIKE」^{☆1}が「ATB」^{☆2}や「RTO」^{☆3}「出発遅延」といった複数の重大トラブルの原因となり得ることが分かります。図-6は、トラブル原因の空港依存性をグラフ化したもので、

☆1 航行中の航空機に鳥が衝突したり、エンジンに吸い込まれたりすることによって、機体の損傷やエンジン停止などの不安全状態となること。
 ☆2 Air Turn Back: 緊急事態の発生により、離陸した飛行機が出発地の空港に引き返すこと。
 ☆3 Reject Take Off: 緊急事態の発生により、離陸を中断すること。

「BIRD STRIKE」が発生する空港に偏りがあることが読み取れます。図-7は「BIRD STRIKE」の事例に特徴的に現れるキーワード間の相関関係をマップ化したもので、「BIRD STRIKE」の発生パターンを俯瞰することができます。これら一連の流れは分析者が行うものであり、計算機はその支援を行います。したがって、分析者がパターンや傾向を把握・発見しやすいように情報を「可視化」する技術が重要となります。

◆リスク評価

すべてのリスク要因を排除し「絶対安全」を確立することは現実的には不可能であり、事後対応的なアプローチも含めたトータルな視点からリスクを管理・コントロールする必要があります。このような観点から重要となるのがリスク評価であり、限られた時間・予算の中で最高の安全状態を確立するために、リスクの発生確率や影響度などを考慮しながら、対策を実施すべきリスクの絞り込みや優先付けを行います。大量の事例をベースに統計的なアプローチでリスク分析を行うことにより、客観的・定量的な評価を行うことが可能となります。

◆対策立案

図-5にもあるように、トラブルの原因と現象（結果）が多対多の対応関係にあり、複数の原因が同一のトラブルを引き起こしたり、単一の原因が複数のトラブルを引き起こしたりします。また、重大トラブルの場合には、その発生に至るまでに複合的な要因・事象の組合せや連鎖が存在し、そのいずれかの段階で適切な対応を実施していれば、最終的なトラブルには至らなかったというケースも多々あります。したがって、効果的な対策を講じるためには、多くの深刻なトラブルに結びつく根本的な要因を発見し、優先的に対処していくといったアプローチが必要です。リスクマイニングの手法により、トラブルに至るまでの事象の連鎖確率を客観的・

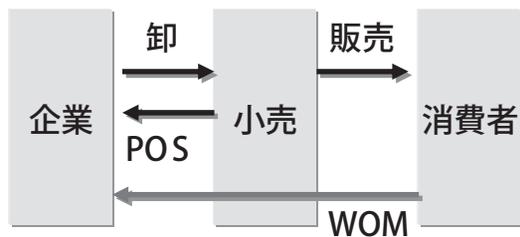


図-11 POSの「量」の分析からWOMの「質」の分析へ

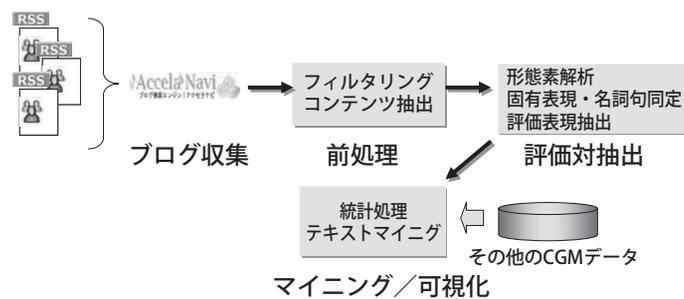


図-12 クチコミ分析システムの概要

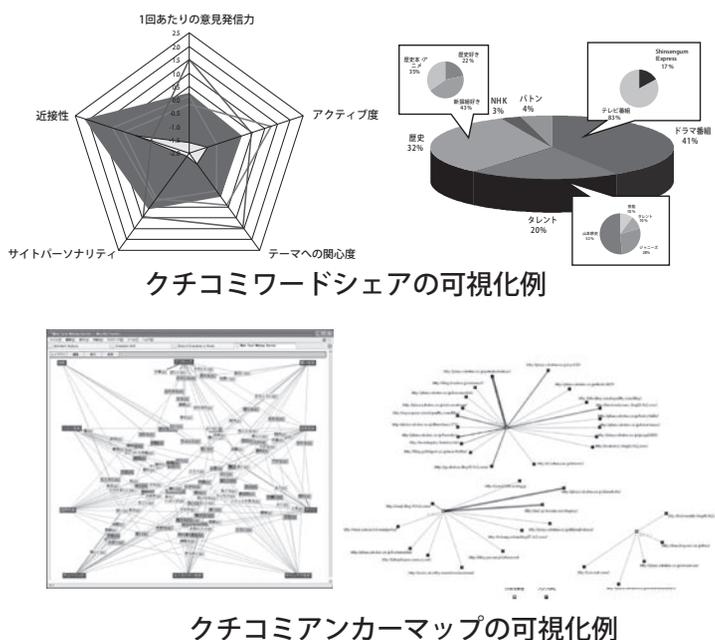


図-13 各メニューの可視化の例

や検索+aの新たなサービスが現れ始めています。WiPi^{☆5}では商品に関するブログのクチコミ情報を検索できます。また、Kizasi^{☆6}や BlogWatcher^{☆7}では、キーワード検索の機能に加えてキーワードのバースト(ブログの記事中で話題になっているキーワード)やキーワードに対する「良い」悪いなどの評判情報を提示することができます。

このようにクチコミに関しては消費者に対してのサービスが多かったのですが、企業向けのサービスとしてニフティの BuzzPulse^{☆8}などが発表されています。これは、ブログを中心とした大量のCGMデータに対し、言語処理とマイニングを組み合わせ、マーケット視点による「見える化」を実現しています。以下では、このCGMのマーケティングへの活

用事例をご紹介します。

◆クチコミ分析システムの概要

一般に公開されているブログ検索エンジン^{☆9}を活用して1億6千万記事(2006年6月24日現在)を対象に、広告記事のフィルタリングやHTMLからの主要テキストの抽出といった前処理を行い、次のステップで言語処理の技術を用いて評価対抽出を行います。さらに、抽出した評価対に対してマイニング処理を施し、さまざまなツールを用いて可視化を行います(図-12)。

◆クチコミ分析のサービス例

現状のクチコミ分析のサービスと可視化の例を以下に紹介します(図-13)。

☆5 <http://wipi.jp/>
 ☆6 <http://kizasi.jp/>
 ☆7 <http://blogwatcher.pi.titech.ac.jp/>
 ☆8 <http://www.nifty.com/buzz/>

☆9 <http://www.accelanavi.com/>

1. 時系列クチコミ評価分析

企業のブランドや商品に関する評価語を抽出し、ポジティブ／ネガティブを判別します。評判の測定や評判に影響を及ぼす要因の特定が可能です。また時系列で比較することで、アクション前後の効果測定にも利用できます。

2. 時系列クチコミ比較分析

企業のブランドや商品に関する評価語を抽出し、ポジティブ／ネガティブを判別します。評判の測定や任意の評価語によって想起されるほかのブランドとの比較が可能です。また、ほかのブランドとの比較を時系列で表示することもできます。

3. クチコミワード 10

企業のブランドや商品と一緒に語られている言葉（キーワード）を、ポジティブ／ネガティブなどに識別し、ランキング表示できます。

4. クチコミワードシェア

ある評価語に対する他社とのシェアをグラフ表示することができます。企業が訴求したいキーワードが他社と比べてどう語られているかといったことが分かります。さらにシェアの時系列変化を見ることも可能です。

5. クチコミアンカーマップ

企業のブランドや商品に関する評価語について、ほかのブランドとともにマッピング表示をすることが可能です。キーワードをビジュアル的に理解できます。

■クチコミ分析の効果

これまでのマーケティング手法と比較して、クチコミ分析を活用したマーケティングの効果は以下のようにまとめられます。

1. 時間：アンケートなどをベースにした従来型の調査では、企画から実施までに1～3カ月を必要としていたが、クチコミ分析では数時間で分析が可能です。
2. 量：従来型の調査では数千件程度のサンプルしか集められなかったが、数万～数百万規模の分析が可能です。
3. 質：従来型の調査では設問や謝礼などによるバイアスがかかっており消費者の生の声を吸い上げることができませんでした。また、ブログなどのCGMを通じた消費者の本音を分析することが可能です。また、Webパーソナリティ（影響力、書き手の属性（性別、年齢））をベースに特定のユーザ群に特化した分析やコミュニケーションが可能です。
4. コスト：従来型の調査では数百万円のコストがかかっていたのに対し、数十分の一以下の大幅なコストダウンが可能です。

CGM マーケティングの今後に向けて

ナレッジマネジメントのマーケティングへの応用としてクチコミ分析の例を紹介しました。大量のCGMを分析して、広告などのプロモーション効果を測定したりや次の製品開発のヒントを見つけたりするプロセスは、まさに分析知を活用するナレッジマネジメントといえます。

CGMを活用したマーケティングは始まったばかりですが、CGMの普及に伴い急速に拡大することが予想されます。今後は、

1. 分析精度の向上：分析辞書やシソーラスを充実させて分析の精度を高める
 2. インフルエンサーの特定：影響力の高い消費者を特定し、その影響度を分析結果に反映する
 3. ブログ以外のCGMの活用：BBS、SNS、ユーザーレビューなどブログ以外のCGMも分析対象とする
- などの課題を中心に研究開発を深耕する予定です。

最後は人間が判断

これまでにご紹介した2つの事例では、ナレッジマネジメントにテキストマイニングの技術を用いることによって、今まではできなかった大量情報の分析が即時にできるようになりました。しかしながら、これらの分析結果を活かして行動につなげていくのは人間です。高いレベルの暗黙知を得ることができるよう、「分析知」を高めていくことが必要です。

参考文献

- 1) 渡部 勇：テキストマイニングの技術と応用、情報の科学と技術、Vol.53, No.1, pp.28-33 (2003).
- 2) 渡部 勇、他：単語の連想関係によるテキストマイニング、情報処理学会 研究報告 FI-55-8 DD-19-8, pp.57-64 (1999).
- 3) 渡部 勇：ビジュアルテキストマイニング、人工知能学会誌、Vol.16, No.2, pp.226-232 (2001).
- 4) 渡部 勇：ビジュアルテキストマイニング技術—特許検索・分析支援への適用—、雑誌 FUJITSU, Vol.56, No.4, pp.371-377 (2005).
- 5) 齊藤孝広、他：障害情報からのマイニング、情報処理学会 研究報告 FI-61-20 NL-142-20, pp.145-152 (2001).
- 6) 齊藤孝広、他：航空安全情報からのトラブル発生パターン抽出について、言語処理学会 第11回年次大会、C1-1, pp.69-72 (2005).
- 7) @IT、野村総研が分析する次世代Webの進化の方向性、<http://www.atmarkit.co.jp/news/200605/19/nri.html> (May 2006).
- 8) 松井くにお、田中穂積：ナレッジマネジメントの新展開、情報処理、Vol.47, No.4, pp.421-424 (Apr. 2006).

(平成18年7月10日受付)

