

「顧客の声」を資産化するテキストマイニング

三室克哉

近年、コールセンターやインターネットを利用した取引により、顧客とのダイレクトなコミュニケーションが実現し、「顧客の声」の蓄積が可能となっている。しかし、これらのデータは膨大で未整理なことが多く、人の手による分析は難しい。そうしたなか、膨大な文章（テキストデータ）を分析する技術として、データマイニングと自然言語処理を融合したテキストマイニングが注目を集めている。この手法を活用して顧客のクレームや意見、要望などを把握できれば、それは企業にとって大きな資産となる。

■「顧客の声」からの発見

従来、卸業者、代理店、営業マンなどを介して、あるいはダイレクトメールなどを通じて伝達されてきた問い合わせやクレームなど「顧客の声」は、今ではコールセンターやホームページなどを通じてダイレクトに企業へと伝えられ、電子的なデータとして蓄積できるようになった。

こうした「顧客の声」を商品開発やサービス改善に活かすことが、企業の生き残り戦略の最重要課題として認識されてきている。しかし、蓄積された膨大な「顧客の声」のデータは、電子化されているとはいえ、文章（テキストデータ）のままであり、人の手によって分析するのは困難である。

そうしたなかで、データマイニ

ング分析技術と自然言語処理技術を融合した「テキストマイニング」が注目を集めている。この分析手法を活用して、蓄積された膨大なテキストデータから、

どの商品に対して何と言っているのか

どんな顧客が何と言っているのか

最近どんな話題が増えているのか（トレンドの変化は？）

優先すべき施策は何か（ニーズの定量的な重要度は？）

施策によってクレームは減ったか

などを発見できれば、それは企業にとって大きな資産となる。

■テキストマイニング

テキストマイニングとは、文字

通りテキストデータをマイニング（採掘）して、ビジネスに役立つ規則、法則を発見する分析手法である。技術的な系統からいえば、テキストマイニングには、新聞記事などの文書検索技術からの発展と、POS（販売時点情報管理）分析などのデータマイニングからの発展の2種類のものがある。

前者は検索や文書分類の効率化を、後者はマーケティングへの活用を目的としている。本稿では特に後者について述べるが、両者とも、きれいに整っていない非定型のデータを扱うこと、かつ、単語ではなく文章（自然言語）を扱うことでは共通している。

（１）データマイニング

本稿で述べるテキストマイニングの基礎となっているデータマイニングは、POSデータによる商品のバスケット分析（変数と変数の相関関係を導き出す手法）、ダイレクトメールでヒット率の高い顧客を抽出するディシジョンツリー分析（要因ごとにYES、NO形式で判定して最終結果を導き出す手法）など、さまざまな分野で活用されて効果をあげている。

通常、データマイニングで使用されるデータは、膨大な量ではあるが、データベースに格納される比較的整理（構造化）されたデー

タ、つまり「項目一値」の固定的なフィールドとして定義されるデータである。そこでは、たとえば単純な「降水量」「気温」ではなく、「雨続き日数」「前日気温差」のように、データマイニングの入力となるデータを事前に加工しておく必要があり、それがプロジェクト成功の鍵でもあった。

しかし、日々変化するビジネス環境のなかでは、定型となったデータはいずれ陳腐化する。本当の発見は、非定型な項目、例えば営業日報の「特記事項」やアンケートの「その他ご意見ご感想」の中にこそ潜んでいることが多い。これらの非定型なデータをいかにして分析の土台に乗せることができるか、テキストマイニングはその解決策の1つである。

(2) テキストマイニングを支える分析技術

テキストマイニングに用いられる手法には、次の4つがある。

文章を単語に分割する形態素解析

「商品」「製品」「品物」といった表記のゆれ、類義語を統一するシソーラス変換（類語・反意語辞典）変換

「商品A」と「汚れる」がいつも同時に出現するといった単語間の関連性を分析する相

関ルール

文章中の単語の係り受け（あ単語が後ろに置かれたどの単語にかかっているか）の関係を抽出する構文解析

文章レベルの分析では、文章の類似性の解析やグループ化を行うニューラルネット（要因ごとに最適な重み係数を求めて最終的な結果を出す手法）、クラスター分析（データを集団に分ける手法）などがある。これらは単語の同時出現に基づくベクトル化など、文章や単語を扱うためのさまざまな工夫がなされている。

テキストマイニングを効果的に活用するには、テキストデータだけでなく、顧客属性、取引情報などのデータと結びつけて分析を行うこと、単発の分析ではなく時系列の変化でとらえることが重要である。また、意味を持った文章を扱うため、「うれしい」「最悪だ」「ひどい」といった良い、悪いとその重要度合いの指標化、要望（欲しい）、疑問（ですか）のような意味的な分類が必要となる。

■ テキストマイニングの活用事例

(1) コールセンターのクレーム分析

蓄積されたクレーム情報をグル

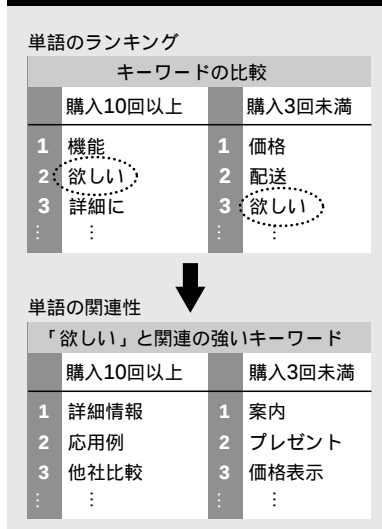
ープ化し、それぞれのグループに含まれるクレームの特徴的な単語（「パソコン」「不親切だ」「配送」など）とそのクレームを発した顧客の属性（性別、取引履歴、地域など）との関連性を見ることが、どんな顧客が何と言っているかが把握できる。また、時系列でクレーム件数の変化を見ることで、クレーム対策の効果を検証することも可能である。

(2) アンケートのフリーコメント分析

ウェブを利用したアンケートなどで、大量に回答が集まり、しかも自由記入欄の記入率が高いと、すべてに目を通すのは困難である。そのような場合にも、テキストマイニングで単純に単語の出現頻度を見るだけでも、全体像をつかむのに役に立つ。

例えば、「(○○して)欲しい」と同時に現れる単語のランキングによって顧客の要望を定量的に順位づけて知ることができ、また「××ですか」と同時に現れる単語のランキングによって顧客が疑問に思うこと、購買の前に知りたいことなどが把握できる。さらに、取引履歴や他の設問とのクロスで見ると、取引の多い顧客とそうでない顧客の意見の比較ができ、施策に優先度をつけることも可能と

図1 単語のランキングと関連性に基づく分析の例



なる。

図1は、顧客の購入経験の差によって意見の傾向がどのように違うかを示したものである。購入経験が多い顧客ほど、詳細情報や他社比較を望んでいることがわかる。

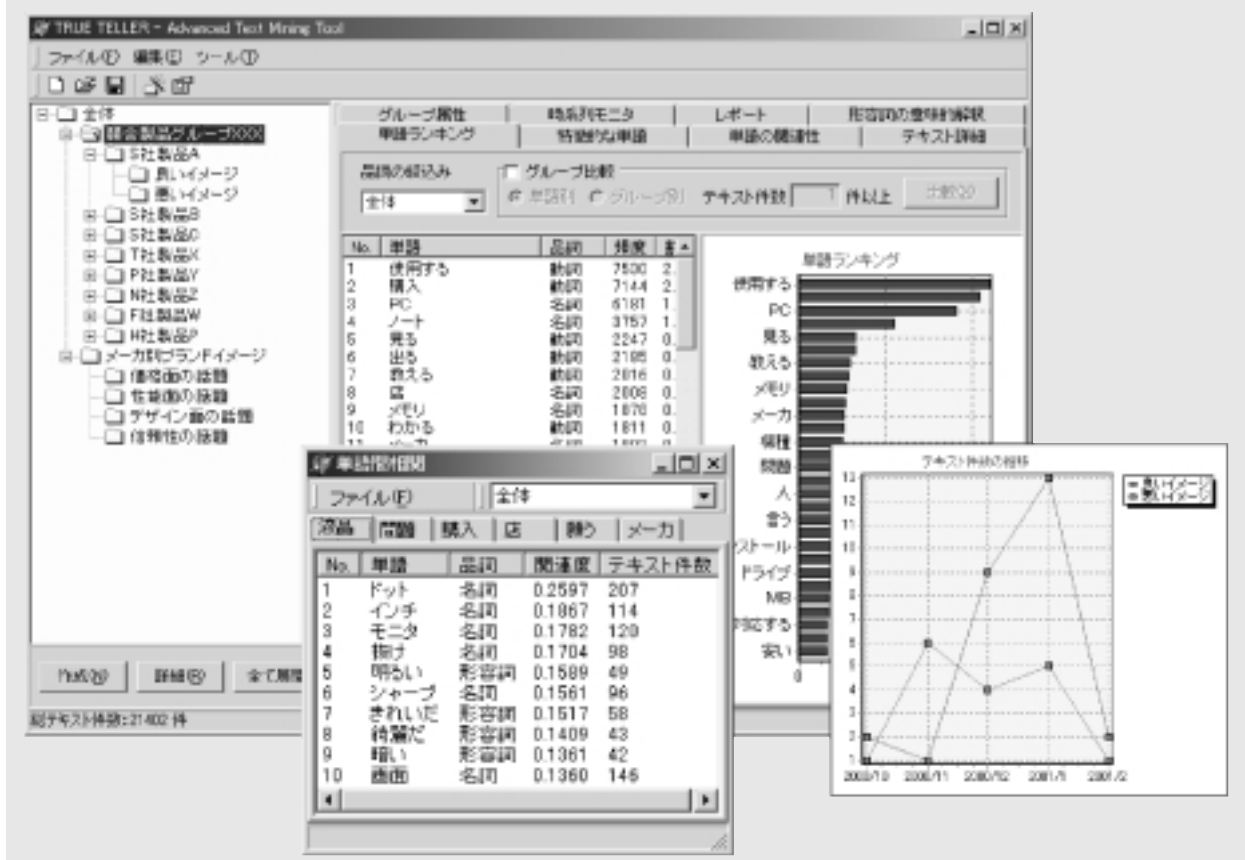
(3) ウェブ掲示板の話題分析、リスク管理

メーカーは通常、自社のコールセンターやホームページから顧客ニーズを収集している。しかし、

故障という具体的な形になる前のものもややした不満や本音の声をとらえることは難しい。

一方、さまざまな掲示板（チャット）サイトでは、消費者による商品やサービスの評価が語られている。メーカーのアンケートと、自由に書き込みされた掲示板の内容を比較すると、後者の方がちょっとした不満から激しい批判まで幅広く、言葉も「最悪な」「ひどい」といった厳しいものが使われ

図2 プロトタイプシステムによる分析画面の例



ていることがわかる。

交通量の多い掲示板サイトを網羅するようにアンテナを張っておくことで、自社のブランドイメージや商品の評価、クレーム、不満などの早期発見が可能となる。

■トータルなソリューションとして

NRI 野村総合研究所では、全体傾向から詳細な文章の分析へと簡単に深化させていける、発見的に思考するようなシステムが必要であると考え、「TrueTeller」(トゥルーテラーと読む。「真実の証人」の意)と名付けたプロトタイプシステムを構築し、分析実験を行っ

た。図2に示したのは、このプロトタイプシステムによる画面の例である。

テキストマイニングツールは、すでにいくつかの製品が販売されている。しかし、「顧客の声」を分析するには、顧客属性、取引履歴、商品情報などと結びつけて分析する必要があり、既存の情報システムとの連携が重要である。

また、ウェブ掲示板の分析のためには、広く外部のウェブサイトアンテナを張りめぐらせ、リアルタイムに自社に関する話題を収集し、商品企画、本質管理に活用することが必要になる。さらに、単発の分析ではなく、継続的に評

価、感想を集め続ける仕組みも必要である。

テキストマイニングはなかなか難しいテーマであるが、蓄積された「顧客の声」という宝の山を、少しずつでも具体的なアクションに結びつけ、活用していくことが重要である。

『システム・マンスリー』

2001年6月号より転載

三室克哉(みむろかつや)

e-システムソリューション部主任システムコンサルタント