

近年、ナレッジマネジメントの重要性が認識されるのに伴い、営業日報、コールセンターへの問合せ、アンケートなど、大量の文書データを自動的に分析し、その内容を迅速に把握してビジネスに生かしたいというニーズが高まっている。そのための自然言語処理技術として、テキストマイニングと呼ばれる研究分野が注目されている。

当社は、このようなニーズにこたえるため、情報抽出に基づくテキストマイニング手法を開発し、この手法を用いて営業日報を分析するシステムを構築した。これは、営業管理職のための意思決定支援を目的として、店頭管理日報からの成功・機会損失事例の分析や、商談管理日報からの受注・失注要因の分析を行うものである。このシステムを利用することによって、膨大な日報すべてに目を通さなくても、時間を掛けずに必要な情報にアクセスでき、営業管理職向け意思決定支援情報としての営業日報の活用が可能になる。

There has been increasing demand in recent years for automatic analysis of large quantities of text data such as salespeople's daily reports, questionnaires, and call center inquiries. Text mining is one of the natural language processing technologies that is attracting attention as a promising solution. To meet the above requirements, we have proposed a method of text mining based on information extraction, and developed a system for analysis of salespeople's daily reports to support decision-making by sales managers using this method. It analyzes "best practice" cases and "missed opportunity" cases from daily reports of storefront management, as well as key success factors and unsuccessful factors from daily reports of business negotiation management.

This system not only enables important information to be quickly accessed without the need to read large volumes of reports, but also makes daily reports a resource to support decision-making.

## 1 まえがき

テキストマイニングは、データマイニングから派生した研究分野である。データマイニングが、関係データベースに格納されている情報の分析を対象としているのに対し、テキストマイニングは、特定のデータ構造を持たない通常の文書を対象としている。情報抽出、文書検索、文書分類、文書クラスタリングなどの要素技術を組み合わせて、大量の文書データから新しい知識を獲得することを目指している。

当社は、情報抽出に基づくテキストマイニング手法を開発し、この手法を用いて営業日報を分析するシステムを構築した。以下、2章で情報抽出に基づくテキストマイニングについて、3章でテキストマイニングの応用事例として、2種類の日報分析システムについて述べる。

## 2 情報抽出に基づくテキストマイニング

### 2.1 表現の抽象化の重要性

特定のデータ構造を持たない通常の文書を分析するためには、まず、文書をコンピュータで扱える形のデータに変換する必要がある。そのためには、言語表現からその意味を抽出する必要があるが、現在の自然言語処理技術では、言

語表現の意味を完全に抽出することはできない。その代わりに、形態素解析という単語に分割する処理をして文書からキーワードを抽出し、そのキーワードで文書の内容を近似する。一般的には、キーワードとして名詞や動詞など内容語だけを用い、助詞や助動詞などの機能語は取り除いてしまう。この結果、言語表現から次のようなキーワードが抽出される。

- (1) “シャンプーは売れませんでした” “シャンプー”、“売れる”
- (2) “シャンプーは売っています” “シャンプー”、“売れる”
- (3) “動きが良い” “動く”、“良い”
- (4) “回転が速い” “回転”、“速い”

しかし、このようなキーワードを対象にデータ処理をすると、(1)と(2)のように、売上げの観点からは、区別が重要なデータが同じものとして扱われてしまう。一方、(3)と(4)のように同じ内容が異なる表現で記述されていると、別の物として扱われてしまう。そこで、文書に含まれる事柄を一つ高いレベルに抽象化することにより、意味のある分析が可能になる。精度の高い分析を行う手法として、情報抽出に基づくテキストマイニング手法を開発した。

### 2.2 テキストマイニングの処理の流れ

テキストマイニングの処理の流れを図1に示す。形態素解

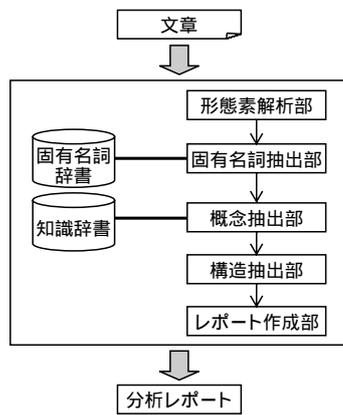


図1. テキストマイニングの処理の流れ メインとなるのは、形態素解析, 固有名词抽出, 概念抽出の三つのプロセスである。  
Process flow of text mining

析, 固有名词抽出, 概念抽出, 構造抽出, レポート作成の五つのプロセスによって, 入力された文書を分析し, レポートを作成する。各プロセスは, 以下のような処理を行う。

- (1) 形態素解析部 入力された文書を単語ごとに分割し, 品詞付けを行う。日本語のように単語ごとに分かち書きをしない言語にとって, 基本となる処理である。
- (2) 固有名词抽出部 固有名词辞書を用いて, 会社名, 製品名, 人名などの固有名词を抽出する。
- (3) 概念抽出部 知識辞書を用いて, 文書中でキーとなる概念(以下, キー概念と呼ぶ)を抽出する。このキー概念は, 概念クラス, カテゴリーに分類しておく。
- (4) 構造抽出部 抽出されたキー概念のうち, 異なるカテゴリーに属する概念を組み合わせ, 構造を抽出する。
- (5) レポート作成部 抽出された概念や構造に基づきデータを集計し, 表やグラフを作成するなど, 分析の目的に応じてレポートを作成する。

### 2.3 知識辞書

知識辞書は, カテゴリー, 概念クラス, キー概念, 表現の4階層で記述する(図2)。各カテゴリーは複数の概念クラス

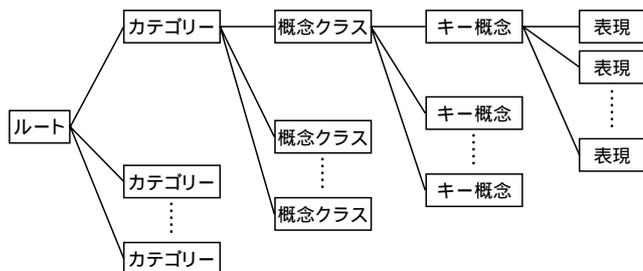


図2. 知識辞書の構造 4階層のツリー構造で記述され, “表現”には文書に出現する様々な文章表現が割り当てられる。  
Hierarchy of knowledge dictionary

を, 各概念クラスは複数のキー概念を持つ。各キー概念は複数の表現に対応している。“表現”は, 文章表現のあいまい性を吸収するため, 正規表現( regular expression )と呼ばれる表記法で記述する。この辞書は, テキストマイニングを適用する目的や文書の内容に応じて用意する。

## 3 テキストマイニングの日報分析への適用

2章で述べたテキストマイニングを日報分析に適用し, 店頭管理日報向けと, 商談管理日報向けの2種類の日報分析システムを開発した。店頭管理日報とは, 各店舗での製品の売行きや陳列状況, 客や店員の評判・対応などを記述した日報である。一方, 商談管理日報とは, システム物件など金額の大きい案件の受注活動に関する日報である。

以下, 両システムの概要について述べる。

### 3.1 店頭管理日報分析システム

3.1.1 システムの概要 このシステムは, ライオン(株)との共同研究によって開発した。日報から成功・機会損失事例を分析することを目的としており, 営業担当者から毎日提出される日報を処理して, 分析結果をレポートにまとめる。このレポートはHTML( HyperText Markup Language )形式で保存され, 本部長, 企画担当者, 営業担当者などは, 各自のパソコン(PC)からWWWブラウザを通じて分析レポートを閲覧する。次の三つの切り口で, 1週間単位で分析を行う。

- (1) 自社情報の分析 自社製品名ごとに, 成功事例か機会損失事例かといった情報の種別と, 店舗業態別に報告件数を整理する。製品名, 店舗業態, 情報種別を指定すると, 指定された情報に関連する日報の要約文と記述文を表示する。
- (2) 競合他社情報の分析 会社名ごとに報告件数を整理する。会社名を指定すると, 指定された会社の製品名ごとに報告件数を整理し, 関連する日報の記述文を表示する。
- (3) エリア別情報の分析 指定されたエリアと店舗業態の店舗名ごとに報告件数を整理し, 関連する日報の記述文を表示する。

画面表示例を図3と図4に示す。図3は自社情報の分析画面である。成功事例は“ベスト(ベストプラクティス)”, 機会損失事例は“ロス(チャンスロス)”という情報種別で表示される。先週と今週の報告件数の増減を色と記号で示すので, 注目すべき情報がひと目でわかる。ここで, 詳しく内容を見たい情報のリンクをクリックすると, 図4に示す記述画面に遷移する。左側に記述文の内容を集約した見出し文, 右側に記述文が表示される。

このように表示することで, 膨大な日報すべてに目を通さなくても, 簡潔に販売傾向を把握できる。数値分析だけの場合と異なり, 文章による微妙なニュアンスや印象を含む定



図3. 店頭管理日報分析システムの画面例(分析画面) 指定された週の自社製品に関する報告の分析である。縦軸は製品名、横軸は店舗形態と情報種別となっており、該当する報告件数が表示される。

Example of system display for daily reports of storefront management: statistics



図4. 店頭管理日報分析システムの画面例(記述画面) 図3で指定した製品名と情報種別に該当する報告の詳細情報である。左側に要約文、右側に実際の記述文が表示される。

Example of system display for daily reports of storefront management: description

性情報も定量的に把握し、経営判断に生かすことが可能になる。また、実際に日報の記述文を読みたい場合にも、機会損失事例を含む日報だけを選択して読むなど、重要な情報を含む日報にすばやくアクセスできる。

3.1.2 分析視点と知識辞書 ここでの分析のポイントは、製品名抽出と成功・機会損失事例抽出である。

- (1) 製品名抽出 対象としたい自社製品名、競合他社名、競合他社製品名を固有名詞辞書に登録しておき、これを参照して製品名抽出を行う。日報では略称や通称も頻出するので、過去の日報を参考にして、正式名称

とともにそれらも辞書に登録する。

- (2) 成功・機会損失事例抽出 店頭管理に関する成功・機会損失事例を抽出するために、販売活動や売上げに関連する概念を扱う。カテゴリとしては、“要因”と“結果”の二つを設ける。要因カテゴリには、販促活動や店頭状況などに関する概念を登録し、結果カテゴリには、売上げや評判などに関する概念を登録する。要因と結果という二つのカテゴリに属する概念を組み合わせることで、“売行き低迷”という結果に関する情報を、なぜ売れないかという要因ごとに分類できる。

知識辞書の記述例を表1に示す。例えば、“販促”に関するキー概念として“景品を付ける”や“POP(Point Of Purchase)を付ける”を登録し、“売上げ”に関するキー概念として“売行き良好”や“売行き低迷”を登録する。これらの概念には、表1に示すような表現が割り当てられている。表現中の“.”、“\*”、“( )”、“[ ]”などの記号は、正規表現で特別な意味を持つ超文字(metacharacter)と呼ばれるものである。

表1. 店頭管理日報向けの知識辞書の例  
Example of knowledge dictionary for daily reports of storefront management

概念クラス	キー概念	表現
販促	景品を付ける	景品.[付つ]ける
販促	景品を付ける	景品.[添貼]付
販促	POPを付ける	POP.[付つ]ける
売上げ	売行き良好	動き.[良好]い
売上げ	売行き良好	売れてい(るます)
売上げ	売行き低迷	動き.(ないありません)

説明のために、語の品詞情報を省略している。  
.: 任意の1文字に適合する。  
[ ABC ]: A, B, Cのいずれかに適合する。  
A B: A又はBに適合する。  
( r ): rが適合するのと同じ文字列に適合する。

このシステムで用いる知識辞書には、13種類の概念クラス、約300個のキー概念が定義され、それに対応する表現が、約1,000個登録されている。

### 3.2 商談管理日報分析システム

3.2.1 システムの概要 このシステムは、当社の東京支店との共同研究によって開発した。案件として継続的に行われている営業活動の日報が対象であり、受失注の要因と競合他社名に着目して分析を行う。競合先ごと、担当部門ごとなどの受失注要因の傾向を分析し、新規案件の営業活動に生かすのが目的である。

東京支店では、従来から当社既存製品である“営業支援システム for Notes”で日報が管理されており、客先別/案件別/担当者別/受失注結果別など、様々な属性によって整

理されている。このシステムは、上記システムと連係して動作し、受失注の要因をこれらの属性や競合他社名と関連づけて提示する。次の三つの切り口で分析を行う。

- (1) 担当部門ごとの受失注要因の統計 営業部門ごとに、受失注要因の構成比率を表形式とグラフで表示する。表の各セルをクリックすると、その部門が担当した案件のうち、対応する要因が記述された日報を含む案件の一覧に遷移する。
- (2) 競合先ごとの受失注要因の統計 競合先ごとに、受失注要因の構成比率を表形式とグラフで表示する。表の各セルをクリックすると、その相手先と競合のあった案件のうち、対応する要因が記述された日報を含む案件の一覧に遷移する。
- (3) 案件の一覧 各案件の日報に含まれる受失注要因に関する記述の計数とともに、案件名を一覧表示する。案件名をクリックすると、その案件に関する日報の一覧に遷移する。

画面表示例を図5に示す。これらはすべて、受注案件と失注案件を区別して閲覧することもできる。

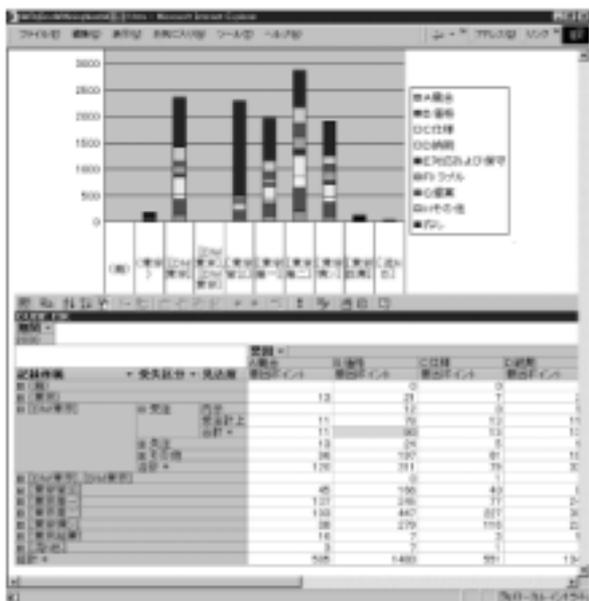


図5. 商談管理日報分析システムの画面例 競合, 価格, 仕様, 納期, 対応及び保守, トラブル, 提案, その他, なし, の九つの要因別に集計される。

Example of system display for daily reports of business negotiation management

3.2.2 分析視点と知識辞書 ここでの分析のポイントは、競合他社名抽出と受失注要因抽出である。

- (1) 競合他社名抽出 対象としたい競合他社名を固有名詞辞書に登録しておき、これを参照して競合他社名抽出を行う。

表2. 商談管理日報向けの知識辞書の例

Example of knowledge dictionary for daily reports of business negotiation management

概念クラス	キー概念	表現
要因	納期	即納.*強
要因	納期	納期.*[ま間][あ合]わ
要因	価格	価格.*好評
要因	価格	価格.*下げられない
要因	仕様	スピード.*好評価
要因	仕様	スペック.*満た.*ない

説明のために、語の品詞情報を省略している。  
.: 任意の1文字に適合する。 A\*: 0個以上のAに適合する。  
[ABC]: A, B, Cのいずれかに適合する。

- (2) 受失注要因抽出 受失注に関連する要因として、“納期”、“価格”、“仕様”といった概念を扱う。表2に知識辞書の記述例を示す。例えば、日報に“即納はやはり強かった”という記述がある場合、表2の1行目の正規表現と適合し、“納期”という受失注要因が抽出される。

このシステムで用いている知識辞書には、受失注要因として8種類のキー概念が定義され、それに対応する表現が約380個登録されている。

#### 4 あとがき

情報抽出に基づくテキストマイニングの処理概要と知識辞書、及び日報分析への適用例について述べた。

3章1節で述べたシステムは、現在、ASP( Application Service Provider )の形態で“ WHYリサーチ ”サービスとして運用されている<sup>1)</sup>。また、2章で述べたプロセスに関してもエンジンとして整備し、自然言語検索や自動分類及びクラスタリングなどのエンジンと連携したナレッジマネジメントツール“ KnowledgeMeister<sup>TM</sup> ”として製品化されている。

知識辞書は、分析目的や業務の種類に応じて構築する必要がある。今後は、日報以外の文書への適用可能性を探るとともに、知識辞書の効率的構築に関する研究開発を進める。

#### 文献

- (1) 岩田誠司,ほか. 顧客の声( VOC)を活用したマーケティングサービス ( WHYリサーチ ). 東芝レビュー . 55, 12, 2000, p.64 - 67 .



市村 由美 ICHIMURA Yumi

研究開発センター 知識メディアラボラトリー。自然言語処理技術の研究・開発に従事。情報処理学会, 人工知能学会, 言語処理学会会員。  
Knowledge Media Lab.



鈴木 優 SUZUKI Masaru

研究開発センター 知識メディアラボラトリー。情報検索技術の研究・開発に従事。情報処理学会, 人工知能学会会員。  
Knowledge Media Lab.