

# カスタマーセンター支援システム

## Support System for Customer Center

岡本 青史  
Seishi Okamoto

(株) 富士通研究所 コンピュータシステム研究所 知能システム研究部  
Intelligent Systems Laboratory, Computer Systems Laboratories, Fujitsu Laboratories Ltd.  
seishi@flab.fujitsu.co.jp

関口 実  
Minoru Sekiguchi

(同 上)  
minoru@flab.fujitsu.co.jp

三末 和男  
Kazuo Misue

(株) 富士通研究所 コンピュータシステム研究所 ドキュメント処理研究部  
Document Processing Laboratory, Computer Systems Laboratories, Fujitsu Laboratories Ltd.  
misue@flab.fujitsu.co.jp

西野 文人  
Fumihito Nishino

(同 上)  
nisino@flab.fujitsu.co.jp

**Keywords:** customer center, call center, help desk system, electronic mail processing, text mining.

### 1. はじめに

従来、コールセンターは、電話やファクシミリなどを用いた顧客からの問い合わせに対し、製品情報やトラブルの対処法などを提供することで、顧客のサポート業務を行う窓口としての役割を果たしてきた。このコールセンターに、近年、大きな変化が起きている。主な変化のポイントとして、以下の2つが挙げられる。

- 顧客とのコミュニケーションチャネルの拡大
- 企業における位置づけの変化

1つは、インターネットの発展を受けて、コールセンターと顧客とのコミュニケーションチャネルが、従来の電話やファクシミリから、電子メールや Web を含んだチャネルへと拡大してきていることである。例えば、コールセンターへの電話は繋がりにくいことが多いため、電子メールによる顧客からの問い合わせを受け付けることで、問い合わせに要する顧客の拘束時間を短縮させることができるようになってきている。また、コールセンターのオペレータと顧客とが、Web 画面を共有し、静止画や動画などのマルチメディア情報を使用することで、顧客への適切で迅速な対応を行うことが可能となってきている。このように、現在のコールセンターでは、多様なコミュニケーションチャネルへの対応を行うことで、顧客満足度を向上させるサービスの提供を行う必要が出てきている。

もう1つは、各企業におけるコールセンターの位置づけや役割が変化してきていることである。従来、コールセンターは、顧客からの問い合わせへの対応業務を行ってきた。しかしながら、最近では、顧客との対話を通じて、顧客の嗜好や関心を抽出する場としても位置づけられるようになってきている。これらの情報を利用することで、各顧客毎に適切な対応を行うことが可能になるばかりでなく、顧客の嗜好や関心を満たす製品の紹介や販

売をコールセンターで行う役割も期待されるようになってきている。また、顧客からの問い合わせや苦情の中から、自社の製品やサービスの問題点を見つけ出すことで、将来の製品開発やサービスの提供に役立つ情報を収集する場としても位置づけられるようになってきている。このように、従来のコールセンターは、構築や運用にコストだけがかさむコストセンターであったが、新たな利益を生み出すプロフィットセンターとしての役割や業務が期待されている。

上述した多様なチャネルへの対応を行い、コストセンターからプロフィットセンターへと変革を遂げたコールセンターのことを、ここでは、カスタマーセンターと呼ぶことにする。本解説では、カスタマーセンターの構築や運用を支援するために、重要な役割を担う以下のシステムの解説を行う。

- ヘルプデスクシステム (2章)
- 電子メール処理システム (3章)
- テキストマイニングシステム (4章)

各章において、我々が開発したシステムの説明を行うとともに、実際の現場で必要とされている技術や機能を紹介し、今後の課題について述べる。最後に、5章において、本解説のまとめと、カスタマーセンター全体の課題について述べる。

### 2. ヘルプデスクシステム

カスタマーセンターのオペレータは、顧客からの問い合わせに対して、適切で迅速な対応を行わなければならない。しかしながら、最近の製品開発のサイクルは早い場合が多く、オペレータのスキル向上のための教育だけでは、顧客の満足する対応を行うことが難しくなってきている。また、オペレータの熟練度の差によって、顧客

へのサービスのレベルがばらついてしまうことも大きな問題となる。このような背景から、オペレータの顧客対応業務の支援を行うヘルプデスクシステムの開発が、カスタマーセンターにおける重要な案件となっている。

本章では、ヘルプデスクシステム全体の概説を行い、その中でも特に重要な役割を果たす質問応答システムの詳細な説明を行う。また、ヘルプデスクシステムにおける今後の課題を述べる。

## 2.1 ヘルプデスクシステムの概要

電話による顧客からの問い合わせに対して、オペレータの対応支援を行うヘルプデスクシステムの画面を図 1 に示す。以下、図 1 を参照しながら、ヘルプデスクシステムが提供している支援機能の説明を行う。

顧客からの電話は、まず、CTI (Computer Telephony Integration) により処理される。発信者番号通知機能が使用できる場合は、顧客 DB 中の電話番号とのマッチングを行うことで、顧客情報を自動的に表示することが可能となる。また、顧客の名前や ID をキーにして、顧客情報を検索できる機能も提供されている。これらの機能により、対応の迅速化を図ることができる。

顧客への実際の対応は、スクリプト化された対応の順番に沿って行われる。図 1 では、右下のフレームが、この処理手順を示している。これにより、初心者オペレータでも適切で迅速な対応が可能となり、オペレータの熟練度の差によって、顧客へのサービスのレベルがばらついてしまうことも防ぐことができる。なお、スクリプト中の「継続問合せ」をクリックすることにより、過去の問い合わせと対応の履歴を参照することができ、顧客への対応状況を把握することができる。これにより、以前とは異なるオペレータが問い合わせを受けたとしても、適切な対応が可能となる。また、スクリプト中の「FAQ」をクリックすることで、後述する質問応答システムを用いて、製品トラブル等に関する Q&A 事例を検索することができる。

図 1 の左下のフレームには、問い合わせ商品の画面イメージが表示される。例えば、ボタン「背面」を押すと、図 2 に示すような、商品の詳細な背面イメージを見ることができる。このように、問い合わせ商品の画面イメージを見ることで、その商品に関する知識のないオペレータでも、顧客との適切なコミュニケーションを行うことが可能となる。

## 2.2 質問応答システム

質問応答システムは、想定される顧客からの質問とその模範回答のペアである Q&A 事例を蓄積しており、現在の顧客からの問い合わせをオペレータに入力させることで、現在の問い合わせに関連した Q&A 事例の検索を行うシステムである。現在の顧客からの問い合わせに対する回答となる Q&A 事例を検索することができれば、

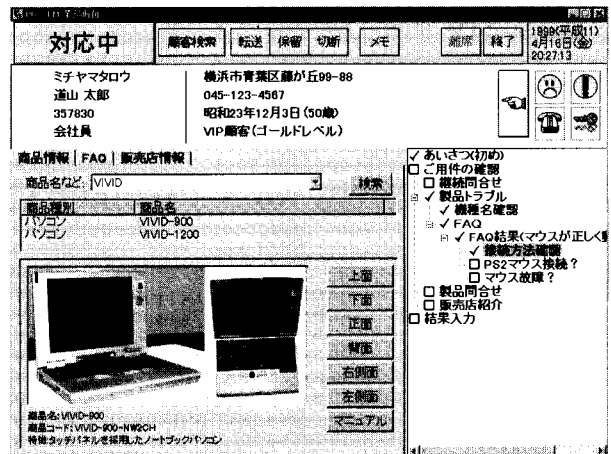


図 1 ヘルプデスクシステムの画面

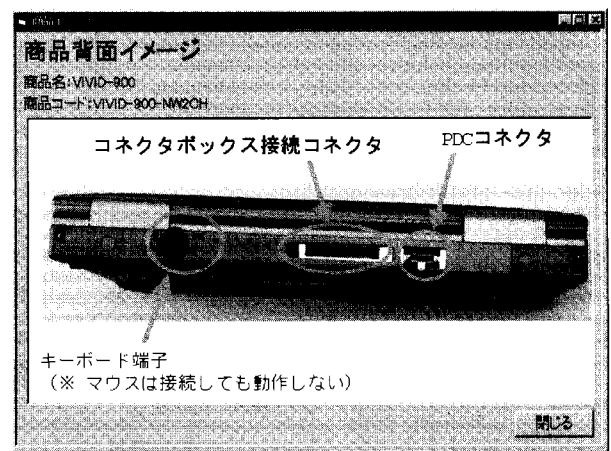


図 2 詳細な商品イメージ

オペレータは、その事例を参照するだけで、顧客への適切な対応を行うことができるため、質問応答システムは重要なオペレータ支援システムとして位置づけられており、人工知能の分野においても、数多くの研究が行われている (例えば、[Hammond 95, Shimazu 94])。

以下では、我々が開発した質問応答システムの検索機能、運用支援機能を説明し、さらに、検索精度に関する評価実験の結果を示す。

### §1 検索機能

我々の質問応答システムは、ランキング検索とディレクトリ検索の 2 つの検索機能を提供している。

ランキング検索は、現在の顧客からの問い合わせを、自然言語文あるいはキーワードのブール式でオペレータに入力させることで、入力との類似度の高い順番に Q&A 事例のランキングを行う検索機能である。ランキング検索は、キーワードのベクトル空間法に基づいて、実現されており、形態素解析ツール [颯々野 97] を用いて、Q&A 事例や質問文中から自動的に抽出された単語や複合語をキーワードとしている。ここで、どのような単語や複合語をキーワードとするかは、システム運用者がカスタマ

イズできるようにしてある。例えば、名詞や形容動詞のみをキーワードとして抽出することもできるし、形容詞や副詞をキーワードとすることもできる。

ランキング検索では、ランキング精度の向上が、オペレータの作業効率の向上の上で重要となる。ランキング精度の向上のためには、キーワードの出現頻度に関する統計量である TF(Text Frequency) や IDF(Invert Document Frequency) に基づいた手法が良く使用される [Salton 89]。これに加えて、カスタマーセンターでは、製品カテゴリやトラブルの種別に応じて、Q&A 事例がカテゴリ別に格納されている場合が多いため、このカテゴリ情報を一種の領域知識と見なして利用することで、ランキング検索の精度の向上が期待できる [岡本 95]。我々は、カテゴリ情報を利用した属性の重み学習手法である CCFI (Cross Category Feature Importance) [Creecy 92] と、TF/IDF 法を融合した類似度関数を使用することで、ランキング精度の向上を図っている。

一方、ディレクトリ検索は、顧客からの問い合わせの状況に応じて、画面に表示されるディレクトリを辿っていくことで、適切な Q&A 事例を見つけることを支援する機能である。ここで、表示されるディレクトリは、製品のカテゴリやトラブルの種別に基づいて、作成されている。顧客からの電話を受けたオペレータは、迅速な対応が必要となるため、ランキング検索における質問入力への負担が無視できない場合もある。このような場合、問い合わせの製品や、トラブルの状況に従って、ディレクトリを辿るだけで、所望の Q&A 事例を検索できる可能性があるディレクトリ検索は、オペレータに対する有用な支援機能となっている。

## § 2 運用支援機能

質問応答システムが効率的かつ効果的に運用されるためには、運用者への支援機能も重要な役割を果たす。

カスタマーセンターでは、同じような問い合わせを、1日に何度も受ける場合が珍しくなく、検索サーバを運用したままで、追加や修正された Q&A 事例を検索インデックスに反映させる機能が重要となる。我々の質問応答システムは、指定時間毎に、Q&A 事例の追加や修正、削除を自動的に検出することで、検索インデックスを定期的に更新する機能や、スーパーバイザからの要求に応じて、即座に検索インデックスを更新する機能を備えている。

また、検索の精度を上げるための運用支援も重要となる。特に、我々のシステムでは、形態素解析により、キーワードの抽出を自動的に行っているため、システム運用者の意図とは異なるキーワード抽出が行われる場合もあり、キーワード抽出を制御するための機能が重要となる。そこで、我々のシステムでは、商品名や型名を固有名詞としてユーザ辞書に登録する機能や、表記の揺れや同義語を定義する機能などを提供している。

## § 3 評価実験

実際の現場での有効性を検証するという観点から、質問応答システムのランキング検索の精度に関する評価実験を行った。

検索対象となる訓練事例として、実際に運用中の質問応答システムに格納されている、パソコンに関する Q&A 事例 21,000 件を使用した。また、テスト事例としては、実際にユーザからシステムに寄せられた質問をランダムに選び、その中から、専門家が、質問に対する正解事例が訓練事例中に存在すると判断した 57 件の質問を使用した。ここで、テスト事例となる質問は、通常の記事で記述されており、バージョンの読み換えを除いて、質問に対する完全な回答が記述されている Q&A 事例のみを正解事例とした。

我々の質問応答システムは、57 件の質問のうち、53 件 (93.0%) の質問に対して、正解となる Q&A 事例を 10 位以内に検索することができ、運用上問題のない検索精度を持つことが分かった。また、検索件数を 5 件とした場合でも、51 件 (89.5%) の正解事例を検索することができ、正解事例は、検索結果の上位の方に検索されていることが分かった。

### 2.3 ヘルプデスクシステムの課題

評価実験により、我々が開発した質問応答システムは、質問に対する正解となる Q&A 事例が存在する場合は、運用上問題ない検索精度を持つことを示した。しかしながら、正解となる Q&A 事例が存在しない場合は、当然のことながら、質問に対する正解を提示することはできない。このように、現在の質問応答システムでは、検索精度自体を向上させることを目的とした研究以上に、高品質な事例を高い生産性で作る技術が求められている。しかしながら、従来の人工知能の研究では、事例の自動作成や作成支援に関する研究はあまり行われてこなかったように思える。今後の人工知能研究において、これらを実現する技術の出現が期待されている。

また、質問に対する Q&A 事例を単に検索するのではなく、質問に対する回答を行うための知識ベースを使用したヘルプデスクシステムの研究も行われている。例えば、知識の獲得や更新の容易性を特長とする Ripple Down Rule を用いたヘルプデスクシステム [Kang 97] や、知識ベースを自然言語で記述することを特長とする対話型ヘルプデスクシステム [日笠 99] などが提案されており、実際の現場への適用が期待される。また、1999 年に開催された TREC-8 (The 8th Text Retrieval Conference) でも、新聞記事の中から、質問文に対する回答を含む文字列を取り出すという課題のコンテストが行われた。ここでは、自然言語処理技術、特に情報抽出技術の利用が注目されており [Srihari 99, Voorhees 99]、今後の研究に期待したい。

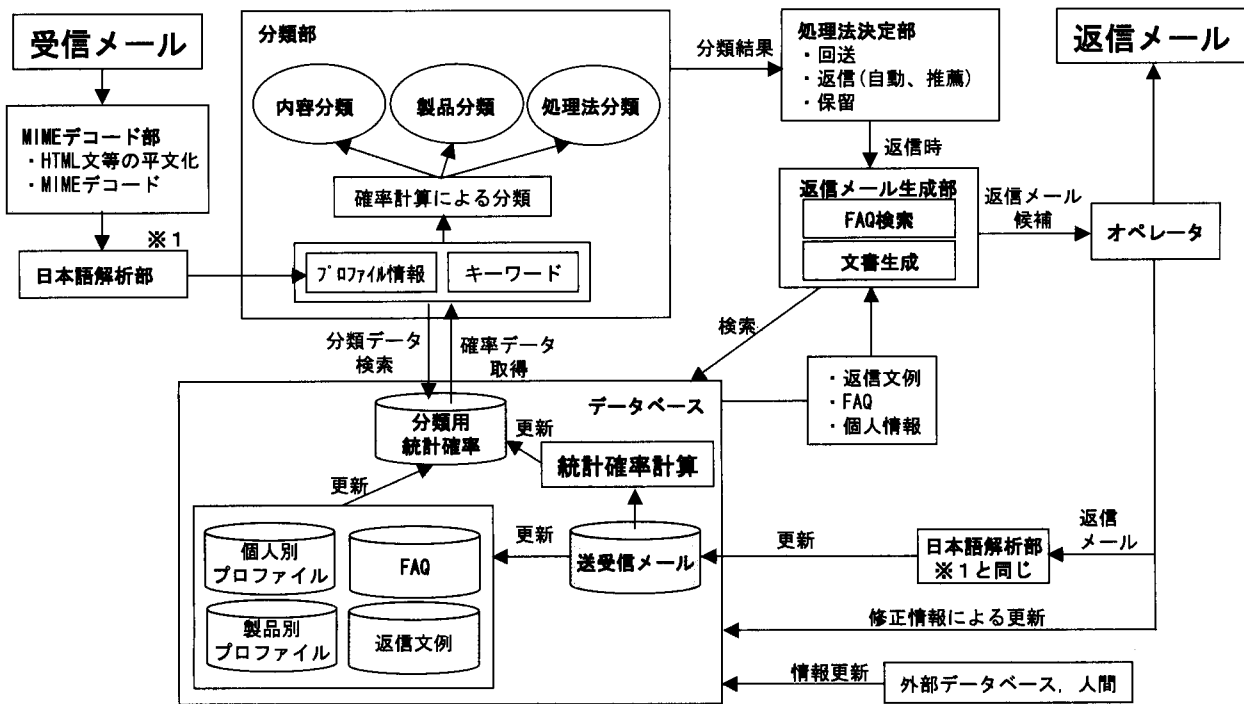


図3 電子メール自動処理システムの構成

### 3. 電子メール処理システム

インターネットの普及に伴って、電子メールで問い合わせやサービスの依頼を行う顧客が増えてきている。カスタマーセンターでも、一般のメーラを使ってそれに対応するだけでなく、専用のシステムを使うことで、運用コストの削減や情報収集の効率化を図ろうとしている。専用の電子メール処理システムには、以下の機能が必要とされている。

- 電子メールのヘッダや本文を分析することで、適切なオペレータを選んで配信するメールホスティング機能
- 定型的な問い合わせに自動で返信する自動回答機能
- 過去の返信文を参照することで、オペレータの返信文作成を支援する返信文作成支援機能
- 高度な知識を有するスーパーバイザへの転送機能
- メール発信者の過去の履歴の参照機能
- 受信・発信メールの管理機能

このような電子メール処理を可能にした、カスタマーセンター向けの市販ツールも欧米を中心に出現しており、日本でも徐々に導入段階にあるが、日本語解析の精度が電子メールの自動応答処理の重要な問題になっている [長谷川 99]。

本章では、我々が開発した電子メール自動処理システムの説明を行い、今後の課題を述べる。

#### 3.1 電子メール自動処理システム

図3は、我々が開発した電子メール自動処理システムの構成、ならびに処理のフローを示している。以下では、

処理のフローに従って、各構成要素における処理内容の説明を行う。

##### §1 MIME デコード部

まず、メールサーバのメーラから取り込んだ受信メールを、MIME デコードや HTML 文の平文化、添付ファイルの分離などのメディア変換処理を施して、メール自動処理システム内での処理が容易な形式に変換する。さらに、電子メールのフォーマット規約に従って、差出人 (From) や受取人 (To) などのプロフィール情報抽出や、サブジェクト (Subject)、添付書類などの抽出を行う。ここで、抽出したプロフィール情報やメールの履歴情報は、個人別プロフィールとして、登録・管理される。

##### §2 日本語解析部

次に、受信メールの日本語解析を行って、メール分類や返信メール作成に必要な情報を抽出する。日本語の解析対象は、差出人名、サブジェクト、本文、署名などである。

日本語解析では、一般に形態素解析によるキーワード抽出が行われるが、ある程度の内容把握ができる日本語解析手法を用いる方が、分類や返信メール作成などの後々の処理のためには都合が良い。このため、日本語解析部では、受信メールからキーワードを抽出するばかりでなく、日本語解析用のルールなどを使用することで、以下の情報の抽出も行っている。

- 文章の構造情報 (挨拶部、本文部、署名部など)
- 固有名の情報 (製品名、会社名、人名など)
- 日付け情報

### § 3 分類部/処理法決定部

日本語解析によって抽出された情報を使って、受信メールをあらかじめ決められたカテゴリに分類する。電子メール自動処理システムでは、受信メールに対して、以下の3種類の分類を行っている。

- 製品分類
- 内容分類
- 処理法分類

製品分類では、テレビ、ビデオ、コンピュータといった製品ごとの分類を行う。内容分類では、受信メールの内容に応じて、質問、苦情、依頼などの分類を行う。処理法分類では、例えば、ユーザ登録の依頼やメーリングリストからの削除といった自動処理や返信文作成処理等の各種処理法に応じた分類を行う。

日本語解析によって抽出されたキーワードを使った受信メールのカテゴリ分類は、Naive Bayes 法 [Duda 73] に、キーワードの出現位置 (サブジェクトや本文など) による重みを導入することで実現している。以下に、分類手法の概略を示す。

分類の対象となるカテゴリを  $C_i (i = 1, \dots, n)$  で表し、受信メールに出現するキーワードを  $k_j (j = 1, \dots, m)$  で表す。また、キーワード  $k_j$  の出現位置に対する重みを  $W(k_j)$  とする。この時、以下で定義される  $S(C_i)$  の値が最大となるカテゴリにメールを分類している。

$$S(C_i) = P(C_i) \prod_{j=1}^m W(k_j) P(k_j | C_i) \quad (1)$$

ここで、 $P(C_i)$  と  $P(k_j | C_i)$  は、過去のメールデータから、統計的に計算される。

この分類処理を、製品分類、内容分類、処理法分類の各々について行い、それぞれの分類に対して、顧客から送られてきた電子メールがどのカテゴリ (例えば、「テレビ」に関する「質問」で「返信必要」など) に属するかを決定する。また、カタログ希望などの典型的な内容のメールに対しては、自動的にカタログ配送の手続きを行うこともできる。

### § 4 返信メール生成部

分類の結果、返信が必要な電子メールに対しては、そのメール内容に対する返信文を作成する必要がある。返信メール生成部では、顧客からの受信メールの日本語解析、および分類結果に基づいて、自動的に返信文の作成を行い、オペレータに提示する。ここで、返信文の中の回答は、受信メールの日本語解析の結果に基づいて、FAQ データベースから適切な回答を検索することで、自動的に埋め込んでいる。また、内容に応じて返信文のテンプレートを選び、それに顧客名や検索結果、署名などはめ込みを行う。これらにより、オペレータが返信メールを作成する作業が、大幅に軽減される。

図 4 は、オペレータの返信処理を支援する処理画面を示している。画面中の返信メールのヘッダや本文などの

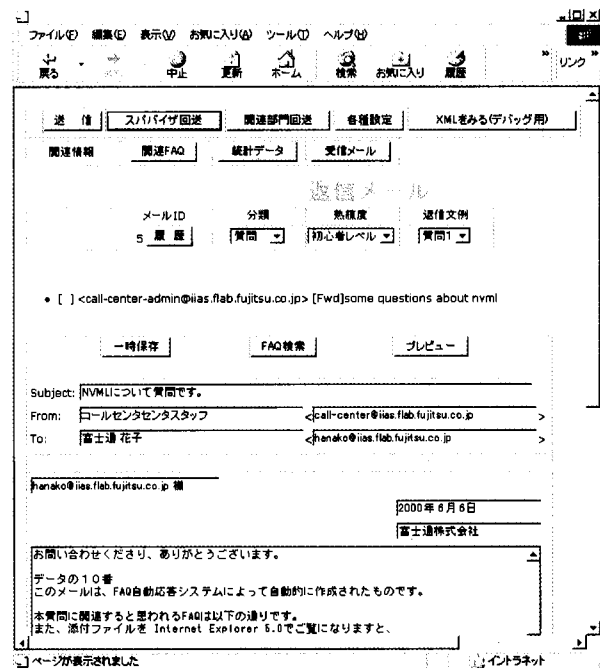


図 4 返信メール作成支援画面

情報は、返信メール生成部において、自動的に埋め込まれたものである。オペレータは、この処理画面に表示されている返信文のチェックや修正を行うことで、実際の返信処理を行う。

### § 5 データベース更新

返信したメールに対し、再度日本語解析が施され、オペレータによる分類誤りの修正や、メールに含まれるキーワード情報などから、分類用の統計確率を更新する。また、うまくヒットしなかった FAQ の追加、修正なども、受信メールと返信メールに含まれる質問と回答から抽出することによって可能となる。

### 3.2 電子メール処理システムの課題

今後の課題として、返信メール作成支援機能の性能の向上が挙げられる。カスタマーセンターに寄せられる電子メールの数は、今後、ますます増加していくことは間違いなく、日本語解析部における情報抽出技術や、回答を埋め込むための FAQ 検索機能などを強化することで、返信メール作成に要するオペレータの負荷の更なる軽減を実現させていきたい。

また、電子メール分類の性能向上も今後の課題として挙げられる。近年、電子メールの分類に関して、様々な手法が提案されており [Brutlag 00, Segal 99]、カスタマーセンターでのメール分類において、これらの手法が、我々の採用した Naive Bayes に基づく手法と比較して、どの程度精度を向上させることができるのかを検証していきたい。

## 4. テキストマイニングシステム

テキストマイニングは、膨大なデータから知識を発掘する「データマイニング」に対して、知識の発掘対象をいわゆるデータ(DB中の数値データや定型データ)からテキストへと換えたものである。テキストマイニングにはさまざまなとらえ方があるが、Hearstは、以下の定義を与えている[Hearst 99]:

真のテキストマイニングとは、テキストデータから貴重な情報(Nuggets)を見つけること。そして、その貴重な情報は、新奇(Novel)のもの。

ビジネス上の情報の70~80%はテキストだと言われている。また、数値データでは言い表わせない微妙なニュアンスはやはりテキストで表現せざるを得ない。ビジネスや生産の現場において、そのようなテキストの分析技術に期待が寄せられるのはごく自然な流れだと考えられる。

本章では、テキスト分析技術のカスタマーセンター支援への適用について述べる。そして、その中心的な役割を担うものとしてテキストマイニングツールを取り上げる。まず、カスタマーセンターにおいて、テキスト分析にどのような期待が寄せられているかを述べる。さらに、その期待に応えるテキストマイニングツールの紹介と、その適用例を説明する。最後に、今後の課題について述べる。

### 4.1 カスタマーセンターにおける期待

カスタマーセンターで扱う情報は大きく2種類に分けることができる。

1種類は、下のようないわゆる「顧客情報」であり、数値データあるいは定型データとして取り扱うことができる:

氏名、年齢、性別、住所、電話番号、職業、家族構成、商品購入動機(選択式)、...

もう1種類は、質問、苦情、要望、ほめといった、いわゆる顧客の「声」であり、一般的には数値データや定型データでは表現できず、テキストデータで表現される。

カスタマーセンターでは、オペレータが入力する顧客との対応ログや、顧客から寄せられる電子メールなどは、このよう顧客の声を反映したテキストデータとして格納されている。そこで、例えば、これらのテキストデータ中の質問に基づいて、FAQをうまく作成することができれば、FAQの作成を行っているスーパーバイザへの支援が可能になる。また、苦情や要望は、短期的には、該当製品やサービスの改善に利用できるし、中長期的には、将来の製品やサービスの企画・開発に利用できる。さらに、ほめの背景を抽出できれば、顧客満足度向上の成功要因として共有できる。

質問傾向分析と新商品企画に対する詳細な要望は、以下のようにまとめられる。

**質問傾向分析** 1. 主要な問い合わせの把握, 2. その原因の究明, 3. 効果的な対策の立案, 4. 対策の効果(問い合わせ状況の変化)の調査

**新商品企画** 1. 要望の整理, および重要な要望の把握, 2. 要望の背景/本質の究明, 3. 本質的に求められている商品の企画, 4. 新商品に対する反応の調査と, 調査結果の次の企画への活用

これらの要望に対し、数値データや定型データだけで、必要な情報を十分に得ることは困難である。例えば、質問傾向分析において、問い合わせ件数や時間の変化は数値データで把握できるが、その背景までは説明できない。また、定型データとして扱えるよう質問項目をあらかじめ選択式にしたのでは、新しい質問傾向に対応できない。新商品企画についても同様である。あらかじめ要望を予想して定型データとして整理したのでは、予想しなかった新しい要望を見失うことになる可能性がある。

カスタマーセンターにおいて行いたいことは、大別すると、傾向・変化・差異の把握や、原因・背景の把握、将来の予測であるが、いずれにおいても非定型なテキストデータを対象として、有用な情報を発掘することが期待されている。

### 4.2 テキストマイニングツール: ACCENT

「テキストマイニング」と称したツールはいくつか開発されているが、いわゆる情報検索や文書分類を越えたツールはまだあまり多くない。

カスタマーセンターに適用されたテキストマイニングツールとしては、TAKMIがある[那須川 99]。TAKMIは、顧客からの問い合わせに関する報告書を対象として、概念抽出の機能を備えており、抽出された概念の量をカテゴリ別や時系列で見せることが可能である。

我々は、人間の「発見」を助けるための有効な手段である可視化技術に着目して、テキストマイニングツールACCENTを開発した[渡部 99]。ACCENTは、文書中から抽出した概念間の関係を図解する可視化機能を有している。この図を通して、概念間の関係の特徴的な部分や変化の傾向を、視覚的に発見・把握することができる。

以下、ACCENTの適用例を紹介する。ただし、この分野の情報はほとんどが企業の貴重な資産であるため、そのまま公開はできない。そのため、ここでは、いくつかの実例を参考にして、人工的に作成したデータを利用してツールの使用例を示す。

対象データは、大規模小売店に電話やハガキで寄せられた「顧客の声」という想定で作成した。「声」の部分は、それぞれ100文字程度のテキストデータである。それと共に「ほめ」、「苦情」、「要望」のような分類も付与してある。

分析手順の概要は以下の通りである。詳しくは文献[渡部 99]を参照されたい。

1. 連想辞書作成 各テキストから単語切り出し、頻度集

計、複合語の分割処理などを行う。

2. **連想検索** 連想辞書を用いて単語の重要度や単語間の関連度を利用して、重要単語や関連単語の検索を行う。
3. **単語マップ作成** 検索した単語やテキストの相互の関係（関連性）をマップとして表現する。
4. **単語マップ操作** マップから求める情報を読み取るために、目的に応じてマップを操作する。

ここでは、よく利用される2種類の単語マップ操作を紹介する。

第1の操作は、単語の固定である。着目すべき単語（概念）を3角形や4角形の頂点に固定することで、それらの比較を容易にする。

図5に、頂点固定マップ（ほめ、苦情、要望の比較）の例を示す。このマップは、「ほめ」、「苦情」、「要望」のそれぞれに分類されたテキスト群と単語の関係を表わしている。さらに、このマップでは、「▲ほめ」、「▲苦情」、「▲要望」を3角形の頂点に固定することで、分類間の比較を容易にしている。例えば、上部に配置されている「親切」、「明るい」、「きれい」などは「ほめ」に特徴的な単語であり、右下に配置されている「有効期限」、「レシート」、「品質」などは「苦情」に特徴的な単語であることが分かる。逆に、中央付近に配置されている単語は、分類間では特徴が見られないということである。

第2の操作は、関係線（グラフで言うところの「辺」）の間引きである。単語間の数多くの関連性の中から、主要なものだけを残すことで、関係構造の骨格を明らかにする[三末 99]。

図6に、辺間引きマップ（ほめの背景把握）の例を示す。このマップは、「ほめ」と関連の強い単語（「ほめ」に分類されたテキストの特徴的な単語）間の関係を表わしている。さらに、このマップでは、関係線を間引くことで骨格構造を顕在化している。このようなマップでは、重要な単語が放射状の中心（ハブ）になる傾向にある。そのため、ハブとなる単語を拾い読みすることで、関係構造の要点を把握することができる。マップから、ほめの要因としては、豊富な品揃え、商品Lの手軽さ、従業員の接客態度などがありそうだとということが分かる。また、「明るい」のは、店内や店舗イメージであることが推測される。

### 4.3 テキストマイニングの課題

テキストマイニングは、まだ比較的新しい研究分野であり、たくさんの課題が残っている。応用によって課題の重要さにも違いがあるが、例えば、質問傾向分析や新商品企画をターゲットとすると、いかに「質問」や「要望」の傾向および背景を明らかにするかが重要である。ここでは、まず顧客の声を表現した大量テキストから質問や要望を選別することが必要である。これは3章で述べた電子メールの処理法分類と同様の作業であるが、その

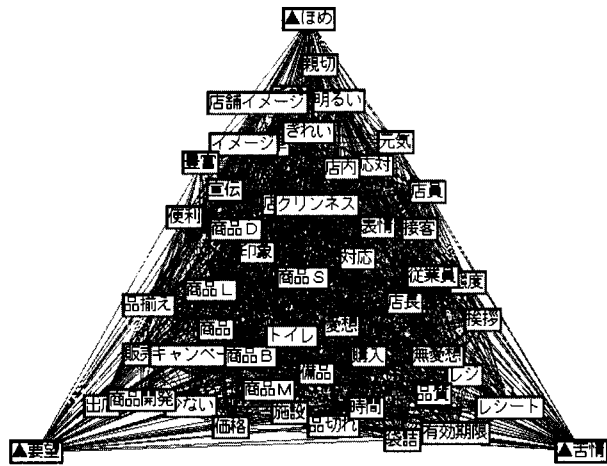


図5 頂点固定マップの例（ほめ、苦情、要望を固定）

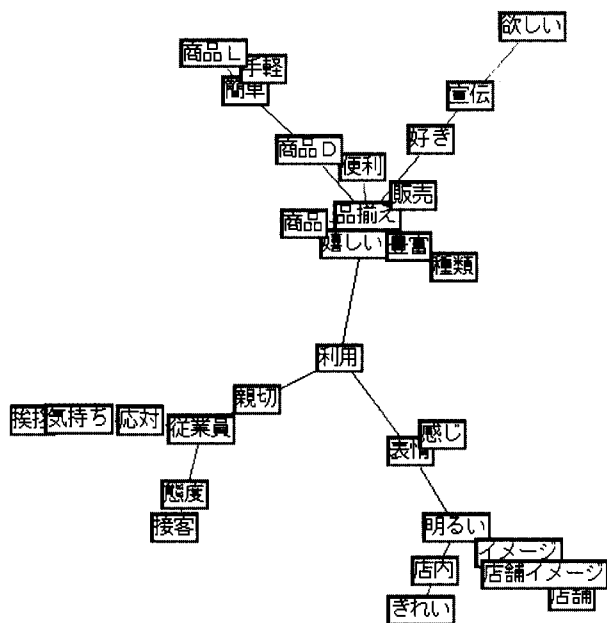


図6 辺間引きマップ（ほめに関連する単語の関連性）

後の分析をできるだけ精密に行うためには、分類のゴミや漏れを極力減らす高精度化が期待される。そうして選別されたテキストから、質問や要望に関わる概念を適切に抽出することも重要である。異表記を吸収し、抽象的過ぎず、特殊過ぎず、適切な単位で概念を切り取ることはそう容易ではない。取り出した概念間に関連付けをすることで概念の全体構造を形成し、可視化することで傾向や背景の把握が可能になる。その関連付けや可視化の手法についても、何が知りたいかという具体的な目的に応じて、さらなる考察が必要である。さらに、そのような構造の比較の簡便化も期待されている。つまり、ある対策前後での質問傾向の変化や、時間経過による要望の変化をも視覚的に概観するための技術である。これにより変化傾向の発見がさらに容易になると期待される。

## 5. おわりに

本解説では、カスタマーセンターの構築や運用を支援する主要なシステムとして、ヘルプデスクシステム、電子メール処理システム、テキストマイニングシステムに着目し、我々の開発したシステムの紹介を行った。また、実際の現場で必要とされている技術や、今後に残された課題について述べた。

カスタマーセンターは、現在、ビジネスの大きな柱として期待されていると同時に、人工知能技術の有効性を実際に示す場としても大きな注目を集めている。本解説で述べた、各システムの現場での課題を解決する人工知能研究の出現に期待したい。

また、現在のカスタマーセンターでは、各コミュニケーションチャンネルに対する個々のシステムが開発されているが、今後は、それらの有機的な連携が重要となってくるであろう。システムの連携や、情報や知識の連携によって、カスタマーセンター全体のパフォーマンスを向上させるための研究にも注目していきたい。

## ◇ 参考文献 ◇

- [Brutlag 00] Brutlag, J. D. and Meek, C.: Challenges of the Email Domain for Text Classification, in *Proc. of the 17th International Conference on Machine Learning*, pp. 103-110 (2000).
- [Creedy 92] Creedy, O. H., Masand, B. M., Smith, S. J., and Waltz, D.: Trading Mips and Memory for Knowledge Engineering, *Communications of ACM*, Vol. 35, pp. 48-63 (1992).
- [Duda 73] Duda, R. O. and Hart, P. E.: *Pattern Classification and Scene Analysis*, New York: John Wiley & Sons (1973).
- [Hammond 95] Hammond, K., Burke, R., Martin, C., and Lytinen, S.: FAQ Finder: A Case-Based Approach to Knowledge Navigation, in *Proc. of IEEE 11th Conference on AI for Applications*, pp. 80-86 (1995).
- [長谷川 99] 長谷川, 湯浅, 堀井: 電子メールのインテリジェントサービス, *人工知能学会誌*, Vol. 14, No. 6, pp. 951-958 (1999).
- [Hearst 99] Hearst, M. A.: Untangling Text Data Mining, in *Proc. of the 37th Annual Meeting of the Association for Computational Linguistics* (1999).
- [日笠 99] 日笠, 藤井, 黒橋: 入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築, *情報処理学会自然言語処理研究会資料*, 99-NL-134, pp. 101-108 (1999).
- [Kang 97] Kang, B. H., Yoshida, K., Motoda, H., and Compton, P.: Help Desk System with Intelligent Interface, *Applied Artificial Intelligence: An International Journal*, Vol. 11, pp. 611-631 (1997).
- [三末 99] 三末, 渡部: テキストマイニングのための連想関係の可視化技術, *情報処理学会情報学基礎研究会資料*, 99-FI-55, pp. 65-72 (1999).
- [那須川 99] 那須川, 諸橋, 長野: 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-, *情報処理*, Vol. 40, No. 4, pp. 358-364 (1999).
- [岡本 95] 岡本, 佐藤, 埴, 松尾: 類似事例検索システム-通信ソフト故障診断問題への適用-, 第 51 回情報処理学会全国大会, pp. 219-220 (1995).
- [Salton 89] Salton, G. and Buckley, C.: Term-weighting approaches in automatic text retrieval, *Information Processing and Management*, Vol. 24, pp. 513-523 (1989).
- [颯々野 97] 颯々野, 斎藤, 松井: アプリケーションのための日

本語形態素解析システム, *言語処理学会第 3 回年次大会*, pp. 441-444 (1997).

- [Segal 99] Segal, R. B. and Kephart, J. O.: MailCat: An Intelligent Assistant for Organizing E-Mail, in *Proc. of the 3rd International Conference on Autonomous Agents*, pp. 276-282 (1999).
- [Shimazu 94] Shimazu, H., Shibata, A., and Nihei, K.: Case-Based Retrieval Interface Adapted to Customer-Initiated Dialogues in Help Desk Operations, in *Proc. of the 12th National Conference on Artificial Intelligence*, pp. 513-518 (1994).
- [Srihari 99] Srihari, R. and Li, W.: Information Extraction Supported Question Answering, (to appear), [http://trec.nist.gov/pubs/trec8/t8\\_proceedings.html](http://trec.nist.gov/pubs/trec8/t8_proceedings.html) (1999).
- [Voorhees 99] Voorhees, E. and Tice, D.: The TREC-8 Question Answering Track Evaluation, (to appear), [http://trec.nist.gov/pubs/trec8/t8\\_proceedings.html](http://trec.nist.gov/pubs/trec8/t8_proceedings.html) (1999).
- [渡部 99] 渡部, 三末: 単語の連想関係によるテキストマイニング, *情報処理学会情報学基礎研究会資料*, 99-FI-55, pp. 57-64 (1999).

2000年9月4日 受理

## 著者紹介



岡本 青史(正会員)

1989年九州大学理学部数学科卒業。1991年九州大学大学院総合理工学研究科情報システム学専攻修了。同年(株)富士通研究所入社。機械学習の理論研究, ならびにその応用研究に従事。1998年博士号(理学)取得。情報処理学会会員。



関口 実

1983年東北大学工学部機械工学科卒業。同年(株)富士通研究所入社。ロボット, 人工知能およびニューラルネットワークの研究に従事。日本機械学会, 日本ロボット学会各会員。



三末 和男(正会員)

1984年東京理科大学理工学部情報科学科卒業。1986年東京理科大学大学院理工学研究科情報科学専攻修士課程修了。同年富士通(株)入社。1990年(株)富士通研究所に転勤。1997年博士号(工学)取得。グラフ自動描画技術およびその発想支援ツールへの応用から始まり, 現在はテキストマイニングおよび Web 文書処理に関する研究に従事。情報処理学会, 日本ソフトウェア科学会, ACM 各会員。



西野 文人

1979年東京工業大学理学部情報科学科卒業。1981年東京工業大学大学院理工学研究科情報科学専攻修士課程修了。同年(株)富士通研究所入社。機械翻訳, 情報抽出など自然言語処理関連技術の研究開発に従事。情報処理学会, 言語処理学会, ACM 各会員。