

永井 明人 高山 泰博 鈴木 克志

三菱電機株式会社 情報技術総合研究所

1. はじめに

Web 文書の検索結果や大量のメール等の大規模な文書から、目的の情報を絞り込んで迅速に入手する目的指向テキストマイニング技術を開発している。本稿では、CRM (Customer Relationship Management) での顧客メールを対象として、CRM 分析者の分析意図に応じたメールを抽出する方式について報告する。

2. メール分析技術の課題

データウェアハウスやコールセンタでは、CRM システムへの顧客メール数が EC 拡大と共に急増している。このため、大量のメールから目的の内容を含むメールを即座に抽出し、クレームへの迅速な対応や、顧客の潜在ニーズ発掘などを実現する要求が高まっており、関連文書検索やメール自動分類の技術が CRM システムへ適用され始めている。ジャストシステムの Concept Base Search [1]等、従来の関連文書検索では単語の統計的な出現傾向のみが類似している文書を検索するため、検索結果に目的外の不要な文書を多く含み、目的の文書の発見が困難である。また、メール自動分類の技術では IBM の意図認識技術 [2]や NTT アドバンステクノロジーの MatchMail [3]等があるが、両者ともメール分類に用いられる知識の作成が容易でなく、システム構築が効率的でない。

3. 提案手法

これらの課題を解決するために、メール文章中の特徴的な言語表現から、分析の観点となる意図（クレーム、要望、賞賛等）を推定してメールを自動分類し、検索時に意図を指定して検索結果を絞り込むことで目的の文書を効率良く発見できるようにする。さらに、メール分類のための知識を自動/半自動的に学習する方式を開発目標とし、種々の適用分野のシステム構築が効率的に行なえることを目指す。

3.1. システムの概要

提案システムの機能構成を図 1 に示す。本システムは、以下の二つの機能により特徴付けられる。

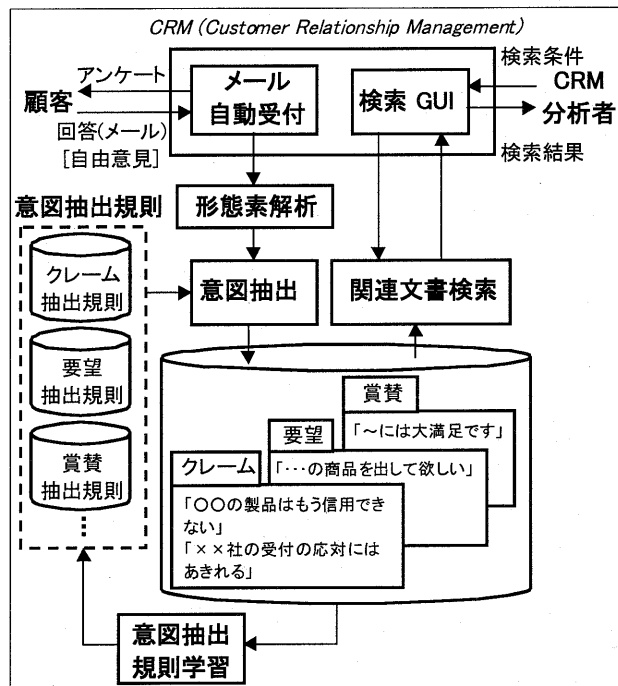


図 1: 顧客メール分析システムの機能構成

- (1) 意図抽出: 顧客のメールから、分析意図に応じた内容のメールを抽出して自動分類する。
- (2) 意図を利用した検索結果の絞り込み: 関連文書検索で得た大量の検索結果に対して、CRM 分析者が指定した意図を含むメールを効率的に絞り込む。

以下、処理フローの概略を説明する。顧客からの問い合わせやアンケート回答等のメールはメール自動受付部により一時蓄積される。意図抽出部は、意図に特徴的な言語表現を単語の共起パターン規則として記述した意図抽出規則を用いて、メール文章の形態素解析結果と照合し、メールの内容に含まれている意図を推定する。さらに推定された意図に応じてメールを自動分類し、データベースへ登録する。分析の際には、まず、分析したい対象を表現する自然言語を入力して関連文書検索を行ない、単語の出現傾向が類似したメールを検索結果として得る。次に、抽出したい意図を指定することにより、検索結果の中に含まれる目的のメールのみが抽出され、顧客データ分析を効率的に行なうことができる。

"A study of a language analyzing method for customer's E-mails of a CRM system"

NAGAI Akito, TAKAYAMA Yasuhiro, SUZUKI Katsushi
Information Technology R&D Center
Mitsubishi Electric Corporation

3.2. 意図抽出方式

意図抽出規則の表現は、コーパスからの自動学習を考慮して、なるべく単純な表現の方が高い学習効果が期待できると考え、表 1 に示すように一文内での単語の共起パタンの重み付き規則で表現する。意図抽出では、入力された文書 D を一文単位に分割し、一文内の形態素列に意図 I の共起パターンが存在すれば、文書 D に対する意図 I のスコアに重みを加算していく。スコアが閾値を越えた意図の中で最大スコアとなる意図 I_{max} を文書 D の推定意図とする。

表 1：意図抽出規則の例

意図	共起パターン	重み
クレーム	・対応 / 悪	1.0
	・納得 / でき / ない	1.0
要望	・し / て / 欲	0.8
賞賛	・とても / 満足	1.0

3.3. 意図抽出規則の学習

意図としてクレームを例に取り、意図抽出規則の学習方式の概要を図 2 に示す。学習データは、抽出意図の有無により、図中(A)、(B)のように二種類に分割する。これらを形態素解析した後に、(A)、(B)共通に出現する自立語を stop word として削除し、出現可能な単語共起パタンの頻度を集計して重みを付与する。これらから、重み値が上位 N 個の単語共起パターンを選択して意図抽出規則とする。

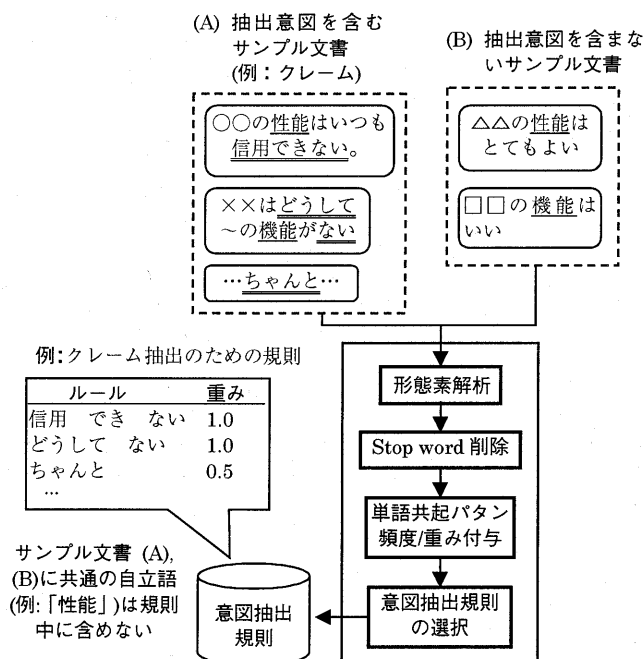


図 2：意図抽出規則の学習方式

4. 初期評価

本方式の意図抽出規則の有効性を予備実験的に評価するため、意図の推定スコアに関して、意図の分離性能を調査した。意図抽出規則は、典型的なクレーム表現の124文を視察して198のクレーム抽出規則を手により作成した。重みは主観的に0.0～1.0の範囲で付与した。評価対象の文書は、上記124文とは別途収集したクレームの文書(110文書)、及び、要望、問い合わせ、賞賛などの文書(101文書)を合わせた211文書である。これらに対し、クレーム抽出規則によるスコア付けを行なった結果を図3に示す。これより、クレーム度のスコアが1.5以上になると、クレーム以外の意図に対して明確な分離性を有していることが分かる。なお、クレームの文書に対するスコアが0の場合が認められるが、これはクレーム抽出規則数が少ないためであり、今後、学習による規則の大規模化を行なって分離性の向上を図っていく。

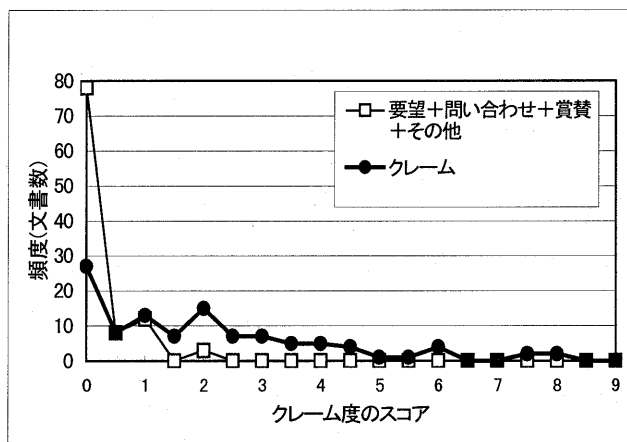


図 3：意図(クレーム)の推定スコアの分布

5. おわりに

CRM 分析者の分析意図に応じた文書(メール)を自動抽出する手法を検討し、初期評価の結果、本手法の有効性を示唆する結果が得られた。今後は、意図抽出規則の学習方式の検討を進め、大規模な文書データに対する性能評価を行なう。

【参考文献】

- [1] 野村, “Concept Base の言語処理と新しいソリューション,” 電子情報通信学会 自然言語処理(NL)研究報告, No. 129, pp. 1-8, 1999.
- [2] 諸橋, 那須川, 長野 “テキストマイニング：膨大な文書データからの知識獲得 - 意図の認識 -,” 情報処理学会 第 57 回(平成 10 年後期) 全国大会 3-75, 1998.
- [3] “日本語完全対応 e メール自動分類・配信ソリューション - MatchMail-CallCenter -,” ビジネスコミュニケーション Vol. 37, No. 5, 2000.