

ユニファイドメッセージを指向するCTIサーバ（3）

5M-8

CTSTAGE向け電子メールの自動内容分類システム

松下 久明 山本秀樹

沖電気工業株式会社 研究開発本部 関西総合研究所

1.はじめに

CTSTAGE[®]は、コンピュータと電話を融合したシステムである。これまでコンピュータと電話で別々に処理してきたメッセージを相互変換することで統一的に扱うことを目的としている。電子メールを電話で聞くために音声変換する場合、単に到着した電子メールを全部読み上げるだけでは冗長であることは否めない。利用者の立場に立つと、システムが電子メールの内容を解析し緊急度の高いメールから読み上げることが要求される。

そこで本論文では、自然言語処理技術によって電子メールの内容を解析し電子メールを分類する機能について述べる。

2.電子メール内容分類機能

2.1.電子メールの構造の解析

電子メールは、RFC1341で定められたようにヘッダ部分とボディ部分からなる。ヘッダ部分の内容はRFCにより内容が規定され構造化されているため計算機処理が容易だが、ボディ部分は自然言語で自由に記述されることが多くこれまで処理対象になりにくかった。電子メールのボディ部分を計算機処理する研究としては、ボディ部分の一部に定型フォーマットを埋め込んだ半構造化メッセージを扱う研究がある（Brownie[1], ObjectLens[2]）。半構造化メッセージは、小規模なグループでは有用だと考えられるが、メッセージ形式を共有しない人々の間では、メールの内容分類には使えない。ヘッダ部分だけから内容を分類しようとすると Subject(件名)フィールド

だけに依存することになり、件名が省かれたメールなどは処理できないので、ボディ部分の解析が必要になる。

ボディ部分は、一般に始めの挨拶、引用、本文、終わりの挨拶といった構造になっていることが多い。内容を分類するためにはまず本文部分と引用部分をうまく抽出し、内容を解析しなければならない。

2.2.電子メールの内容の解析

電子メールの本文で使用される単語はユーザの業種や職務によって変化するが、一人のユーザについて考えると例えば計算機の購買部門では、計算機会社や構成部品名などが頻出すると考えられる。購買部門に必要な電子メールの分類機能は、計算機会社や構成部品名などの分類だけでなく、「連絡」「質問」「回答」など電子メールの発信者の意図による分類も必要である。この電子メールの発信者の意図による分類は、他の業種や職務でも同様に使用可能である。

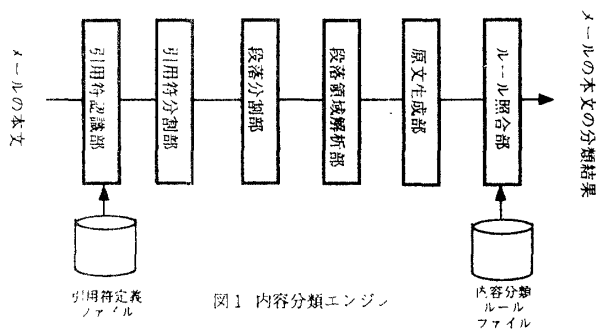
以下では、上記の検討に基づき電子メール分類エンジンの構成と、電子メールを発信者の意図によって分類するためのルールについて述べる。

3.分類エンジンの構成

分類エンジンのシステム構成図を図1に示す。

分類エンジンは、引用符定義ファイルを参照して同一引用符毎に分割する引用符分割部、空行等の段落区切り毎に段落を分割する段落分割部、段落領域部分を解析し、引用符等の余分な文字を省いた文字列を生成する原文生

成部、及び内容分類ルールを参照してルールとの照合を行うルール照合部からなる。



4.分類ルールとルールエディタ

電子メールを「連絡」「質問」「回答」など電子メールの発信者の意図によって分類するためのルールをルール編集用のルールエディタの画面とともに示す。

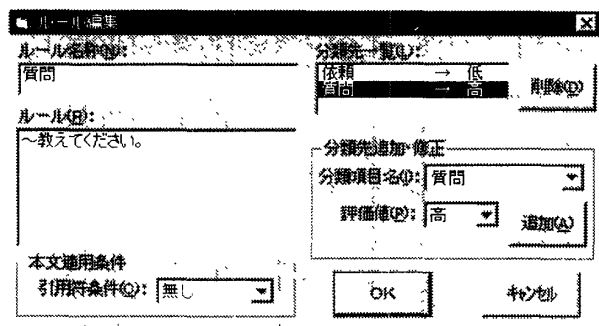


図2 ルール編集画面（質問）

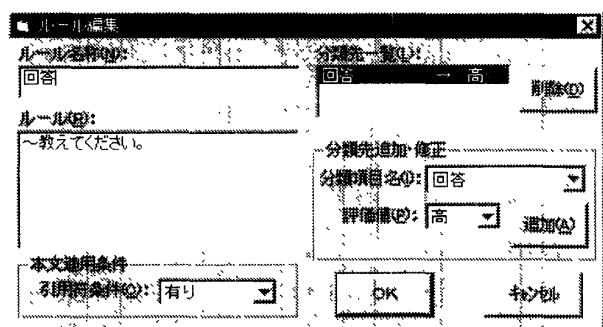


図3 ルール編集画面（回答）

ルールの条件には、ある文字列のパターンが、電子メール中の引用部分か、それ以外のどちらから出現したかを記述する。分類先には、「高」・「中」・「低」の重みのついた複数の分類名を記述できる。図2は、引用符無しで文末に「教えてください。」が出現した時に、

依頼（低）・質問（高）に分類するルールを示している。図3は、引用符有りで文末に「教えてください。」が出現した時に、回答（高）に分類するルールを示している。

5.動作例

分類	優先度	差出人	件名
回答(高)	田中	田中	不明
質問(低)	佐藤	納期確認	
質問(中)	奥村 晃弘	議事録 (12/02)	
指示(高)	管理課	年頭式	
指示(高)	管理課	探しています	
報告(中)	山田	DHQP サーバ	
連絡(高)	postmaster	router stop	
連絡(高)	北 英彦	引越し	
連絡(高)	松下 久 明	ウイルスチェックの実施依頼	
連絡(中)	山本 秀 樹	打ち合せ (11/27)	

図4 内容分類結果画面

図4は、受信フォルダのメールに対して内容分類を行ない分類欄でソートした結果を示している。分類欄は、分類先の評価値を100分率で計算し、最大の分類先及び計算結果を示している。+欄は、差出人毎の重要度を「高」・「標」・「低」で示している。分類欄及び+欄でソートすることにより、単に到着順でなく、緊急度の内容の順に読み出すことができる。

6.まとめ

本論文では、電子メールの内容を分類機能について述べた。本機能は現在開発中である。本機能が搭載されると CTSTAGE は、重要度の高いメールや質問に対する回答だけ優先的に電話で聞くことができるようになる。

今後はより短い時間に電子メールの内容を電話で聞けることが必要になると考えられる。そのため現在、電子メールの要約機能も開発中である。

参考文献

- [1] 奥村晃弘, 北英彦: 非同期分散型グループウェアシエル Brownie の提案, 信学技法 IN91-80(1991).
- [2] Malone, T.W., et.al: Semi-Structured Messages are suprisingly useful for computer-supported coordination, ACM TOIS, pp.115-131(1987)

¹ CTSTAGE は、沖電気工業株式会社の商標です。