1F5-3in

辞書構築及び関心領域設定インタフェースの開発による電子メール文書閲覧支援

Supporting e-Mail Browsing by Developing User Interface for Construction of Keyword Dictionary and Configuration of Interesting Points

*1松本 慎平 Shimpei Matsumoto　*2不動 雄樹 Yuki Fudo　*1秋吉 政徳 Masanori Akiyoshi

*1広島工業大学 情報学部
Faculty of Applied Information Science, Hiroshima Institute of Technology

*2広島工業大学 大学院工学系研究科
Graduate School of Science and Technology, Hiroshima Institute of Technology

Today the Internet is vital to society, and especially e-mail is the most common means of interpersonal communication. Usual e-mail systems have ex-classification function based on folders or tagging, and the function is simple but usable. From now some e-mail browsing support systems with text analysis function have been proposed and developed to improve readability of e-mail text, and excellent outcomes were obtained. However many of them was the system designed to be only e-mail browsing support, and there have been few works developing a function or an interface to customize into user or organization specialized keyword dictionary through e-mail operation. Therefore this paper focuses on the content of a mailing list, and develops user interface to generate user specialized keyword dictionary while supporting e-mail browsing by detecting interesting points.

1. はじめに

情報端末やインターネットの普及に伴い，情報の取得や公開が容易に可能となった．また，人と人とのコミュニケーションも現在インターネットを中心に行われており，その最も一般的な手段は電子メールである．平成 23 年版情報通信白書によると，PC からインターネットを利用する目的が電子メールの送受信であると回答した割合は 46.4%，同様に携帯電話からの場合は 54.5%であると報告されており，両者共にインターネット利用目的の第 1 位にあげられている [総務省 10].

コミュニケーションを電子メールに依存する部分は大きいが，その反面，電子メールの利用目的の多様化や送受信量の増大に伴い，電子メールを効率的に管理する技術が求められるようになった．メール閲覧行動の支援を目的に，メール受信システムに文章解析機能を組み込んだシステム [山口 10] や，閲覧メールの情報を補完する情報を持ったメールの自動検索・提示 [河重 06] などが提案・開発されており，優れた成果が得られている．しかしながら，それらのほとんどは汎用語辞書を前提としたシステムであると共に，メール文書の可読性を高めるための仕組みに留まっている．利用者とシステムとのインタラクションを通じて，個人や組織に特化したシステムにカスタマイズするための機能やインタフェースについてまでは，深く言及されていない．また，メール文書からの知識獲得を視野に入れた文書閲覧支援に関する成果は見当たらない．

2. 目的設定及び提案システムの概要

本研究では，電子メール閲覧を通じて利用者独自の辞書をオンラインで構築するためのインタフェースと，利用者の関心のある領域を強調して表示することに特化したインタフェースを開発することにより，電子メール閲覧を支援する．1. 膨大な量を受信する，2. 閲覧型の電子メールである，3. 受信した情報の一部分を知識として体系的に管理する場合がある，の 3 点を閲覧支援に重要な特徴であると設定し，管理対象の情報源を学会メーリングリストとした．利用者が必要とする情報を膨大な量の中から的確に抽出する機能があれば大変有益であると考えられる．各学会が提供しているメーリングリストの内容は，論文が書籍化された報告や，研究会の開催告知，研究会のプログラムなどの情報が掲載されている．研究会の開催告知やプログラムには多数の専門的な用語が含まれており，利用者にとって関心の高い用語が含まれている場合は，その著者やタイトルなどを後に参考にする機会が多い．以上メールの多くは，一般的にその本文が空白行によりいくつかの部分領域に分けられている．そこで，部分領域ごとに，関心の高い用語を含む部分領域は興味領域，その逆を無関心領域とする．そして，関心領域を強調表示し，逆に無関心領域をデフォルトで折りたたんだ非表示の状態に設定することで，メール本文全体の閲覧効率を向上させる．1. 関心用語を辞書に登録，2. 関心・無関心を利用者が定義，3. 関心・無関心を単純ベイズ分類器で判定，以上のインタラクションをシステムのインタフェースから動的に可能とする．

連絡先: 松本慎平，広島工業大学情報学部知的情報システム学科，〒 731-5193 広島市佐伯区三宅 2-1-1，五日市キャンパス新 4 号館 319 号室，TEL/FAX: 082-921-6924，
E-Mail: s.matsumoto.gk@cc.it-hiroshima.ac.jp

3. 実装及び関心・無関心判定精度の検証

システムのインタフェースを図 1 に示す．右上の領域には受信したメールが設定順に表示されており，下部には選択されたメールの内容を閲覧できるようになっている．左側のフレームには，辞書の編集や不要語リストなどの諸機能を選択できるようになっている．辞書構築画面を図 2 に示す．辞書の構築では，メール本文に対して，利用者が重要語であると判断した用語を手動で辞書に登録するシステムを構築した．文章内の語をドラッグすることで登録ウィンドウが表示されるため，登録する際の利用者の負担を軽減することが可能である．

関心領域では，まずメール文書に対して，空白行を文の区切りと考え，いくつかの部分領域に分ける．部分領域毎に関心の有無により，部分領域内の語と語の出現頻度をそれぞれデータ

図 1: 提案システムのインタフェース

図 2: 辞書構築インタフェース

ベースに記録する．この作業を繰り返すことで，関心領域の設定を行う．以降，メールを部分領域化した際に，データベースから興味のある場合とない場合の語と語の出現頻度を取得し，単純ベイズ分類器によって興味の有無を判定する．判定の結果 “関心無し” が閾値以上であれば，その部分領域を折りたたみ，非表示にする．逆に “関心有り” が閾値以上であれば，その部分領域を強調表示し，それ以外の場合は，そのまま表示する．また，無関心領域の部分領域が連続して出現した場合，連続した部分領域をひとまとめにできる．

本研究は，電子メールの閲覧支援を一次的な目的としながら，利用者とシステムとのインタラクションの繰り返しから得られる独自の用語辞書構築を副次的な目的とする．具体的には，重要語の強調表示や，分類器学習の際に重要語認識させるためには，文書内の重要語とそれ以外の用語を判別する必要がある．判別には形態素解析技術を用いるが，形態素解析には用語辞書データが必要となる．本研究では，個人や組織に特化した利用を想定し，利用者が独自の辞書を容易に構築可能なインタフェースを開発する．この辞書を単一の利用者か，あるいは同じ価値観を持つ利用者同士で共有することで，利用目的に特化した用語辞書が構築される．

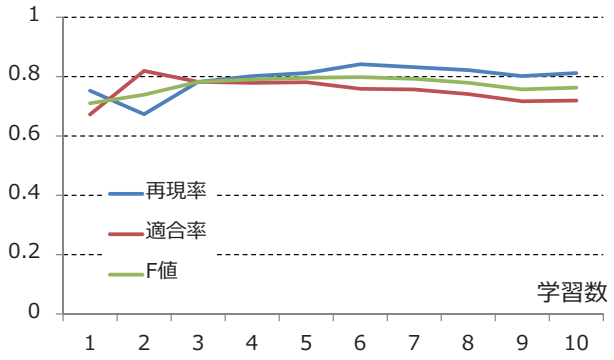

図 3: 関心・無関心分類精度の実験結果

ここで示す実験に先立ち，辞書の精度及び辞書サイズによる処理時間の変化について調査した．実験から，MeCab の形態素解析を利用して用語を抽出する際，全ての用語を包括するような辞書を Wikipedia やはてなキーワードなど既存コーパスから構築するよりも，用途に適した辞書を適切なコストを割り振りながら構築することが，処理時間や精度の観点から有用であった．以上を踏まえ，関心・無関心領域の精度を検証するため，実験を行った．まず，実運用を通じて利用者の嗜好に合った専門用語辞書を構築した．管理されている人工知能学会など情報系関係学会の過去配信メール 3805 件の内，学習用，検証用にそれぞれ 10 件のメールを任意に抽出した．検証用の関心・無関心の正解データを設定し，学習メール数を増やしながら，学習メール件数に対する分類の精度について，適合率・再現率・F 値の評価指標を用いて評価した (図 3 参照)．

4. おわりに

本研究では，電子メールの閲覧支援システムを開発した．また，利用者自身が辞書データをオンラインで対話的に構築できるシステムを構築し，利用者にとって重要な用語が強調表示されるシステムを開発した．膨大な文書全文を読む手間を減らすため，メール本文内をいくつかの部分領域に分け，それらを利用者の関心度の高さに応じて，強調表示や折りたたみ (非表示化) をできるインタフェースを実装した．これにより，利用者は比較的少ない時間で重要な部分のみを確認することが可能となり，文書閲覧を支援した．実験では，任意に選択したメールに対して関心領域の設定を行い，他のメールでも関心領域の適用が行われているかどうかの確認と精度の評価を行った．その結果，小規模な実験ではあるものの，一定の精度が確認された．今後は，分類の精度向上や，得られた利用者独自の辞書の評価方法について検討する必要がある．

参考文献

[河重 06] 河重貴洋，大島裕明，小山聡，田島敬史 他，コンテキストを用いたメールの情報補完，情報処理学会 データベース・システム研究会報告，No.78, pp.329-336 (2006).

[総務省 10] 総務省，情報通信白書平成 23 年版，(2010), http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/, 2013/4/9 参照.

[山口 10] 山口重也, 諏訪敬祐, メール受信における文章解析を活用した情報支援システムの研究, 東京都市大学環境情報学部 情報メディアセンタージャーナル, pp.62-67 (2010).