

## 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

山口 重也 諏訪 敬祐

本論文では、従来のメールシステムで隠されている部分のメールを表示することを目的としている。具体的には、メール受信時にメールに記載されている文章を解析することにより、メールの内容に合わせて配置・装飾をした。動的な配置により、従来のメール受信システムで発生していた新着メール内で重要なメールがその他の新着メールにより隠される新着メール問題が起りにくい表示を実現した。表示結果の変化を調べるために表示状況をシミュレーションし、提案方式が効果を発揮する条件を明らかにした。さらに、実際の環境を想定したメールの送信実験を行い、提案方式の有効性を明らかにした。また、解析結果を踏まえ、メール本文を表示するときにアドバイスする情報支援機能も開発し、動作を確認した。以上の研究により、全体の16%を占める重要分類メールで、分類精度78%、表示件数設定が7件の場合、表示される件数が2.64件(22%表示)から6.84件(58%表示)へ増加し、表示される割合が36%増加することを明らかにした。

キーワード：電子メール、情報推薦、文章解析、フィルタリング、PHP、WEBアプリケーション

## 1 はじめに

### 1.1 研究の背景

現在、電子メールは、様々な場面で利用されている情報通信の手段である。平成21年版情報通信白書によるとパソコンからの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2007年末(平成19年末)で47.2%と、利用目的の第2位に挙げられている。

また、メールマガジンの受信も2008年末(平成20年末)で15.3%となっており、インターネットを利用する目的に電子メールの利用が挙げられる割合が高いことが分かる<sup>[1]</sup>。しかしながら、これらのメールは閲覧のみで利用が終わっているため、記載されている情報資源が有効に活用されていない現状がある。また、重要なメールがその他の新着メールで隠れる新着メール問題(2.1節で後述)が発生しやすい。

メールにおける文章解析の例としては、フィルタリング技術が挙げられる。フィルタリングはスパムフィルタの研究がよく知られている。しかしながら、スパムメール以外の分野に関しては、メールの本文をフィルタリ

ングし、文章解析して情動的価値を評価し、ユーザに合わせた情報を推薦することで新着メール問題を改善する研究はほとんど行われていない。

また、もう一つの背景として、情報推薦技術の発達がある。情報推薦技術とは、ユーザの行動から嗜好を分析、抽出して最適な広告などの情報を推薦表示する技術のことである。メールにおける情報推薦は広告を挿入する仕組みや、内容が類似した過去のメールを参考しやすいように推薦表示する研究<sup>[2]</sup>などがある。しかしながら、メールにおける情報推薦に関する研究は少ない。

そこで、本研究では、電子メールの内容を分析することにより全てのメールの内容に合わせて分類を行い、メールを内容ごとに分類する。そして、メール内容に合わせた情報を推薦する。情報推薦技術を組み込んだメール閲覧、アドバイス表示システムを構築する。このフィルタリングと情報推薦を組み合わせたシステムにより、従来のメールシステムでは表示されない部分のメールを表示する情報支援システムを構築する。

### 1.2 研究の目的

本研究では、前述の背景を踏まえ文章解析を活用したメール受信システムを作成し、ユーザのメール閲覧行動を情動的に支援する情報支援システムを構築する。それにより、推薦設定した分類のメールの表示件数を増加させる。さらにメール本文表示におけるセキュリティ問題を改善したメール受信システムも提案する。また、実証

YAMAGUCHI Shigeoya

武蔵工業大学大学院環境情報学研究科2009年度卒業生

SUWA Keisuke

東京都市大学環境情報学部教授

実験を行うことにより、本システムの有効性を示す。

具体的には、受信したメールを文章解析して記載されている内容を分析し、メールの内容に合わせてメールの配置や装飾を行うメール推薦システムと、メール本文に含まれる文章を解析して内容に合わせたセキュリティ対策やアドバイスを表示するアドバイスシステムを作ることにより、ユーザのメール閲覧行動に対して情報的な支援を行う情報支援システムを構築する。

そして、構築したシステムに対してメールの送信実験などの実証実験を行うことで本システムのメール表示状況を確認し、具体的な改善効果を明らかにする。

## 2 従来のメールシステム

### 2.1 従来の表示方式

メールを送受信するメールソフトウェアにはクライアント型とWEBメール型の2種類が存在している。しかしながら、多くのメールソフト(サービス)では、図1のように受信した電子メールを新着順に表示をする新着リスト方式を導入している。

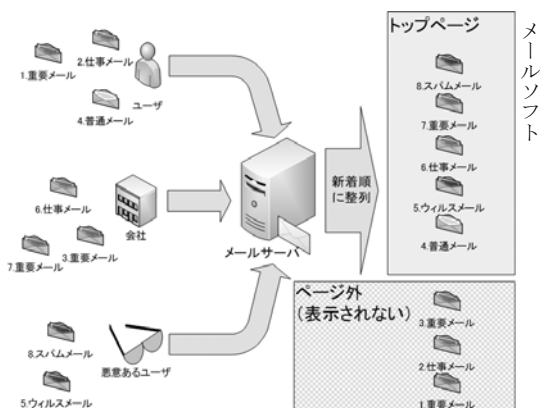


図1 新着メール問題

そのため、最初に送信された電子メールはリスト下部に送られ、最後に送信されたメールはリスト上部に表示される。従って、重要なメールなどが送信された後に迷惑メールや優先度の低いメールが大量に送信されてきた場合、先に送られた重要メールがその他のメールによりページの表示領域外へ隠される新着メール問題が生じる。

### 2.2 従来のフィルタリング

多くのメールソフトでは上記の問題を解決するため、図2のようにフィルタリングを導入している。フィルタリングには2種類あり、1つ目はスパムメールを判別するスパムフィルタである。

スパムフィルタはメーカーが独自にソフトウェアへ組み込んでいるもので、独自のフィルタルールに従って迷惑メールのフィルタリングをしている。スパムフィルタに

迷惑メールであると判断されたメールは迷惑メールフォルダへ配置されるため、受信したメールを表示する画面へ表示されることがない。そのため、スパムフィルタにより迷惑メールをフィルタリングすることで、新着リスト問題を軽減することが可能である。

しかしながら、スパムフィルタは迷惑メール以外のメールには対応していないため、迷惑メール以外のメールが大量に送信された場合は効果を発揮できない。

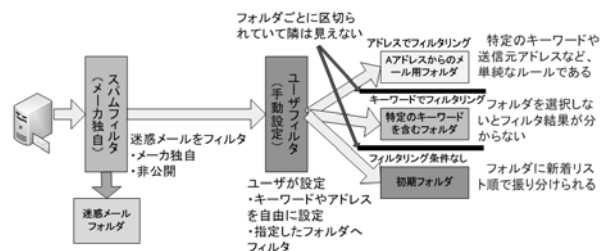


図2 従来のフィルタリング

2つ目はユーザが設定した内容でフィルタするユーザ設定フィルタである。これはユーザが手でフィルタを設定し、ユーザフィルタでフィルタリングしてメールの表示内容を制御する方法である。ユーザ設定フィルタで利用できる条件は、特定のキーワードやアドレス・送受信日時など、限られた条件である。また、設定したキーワードなどを含むメールは内容によらず、同じフォルダに分類される問題がある。

### 2.3 従来の情報推薦

従来の情報推薦の例を図3に示す。

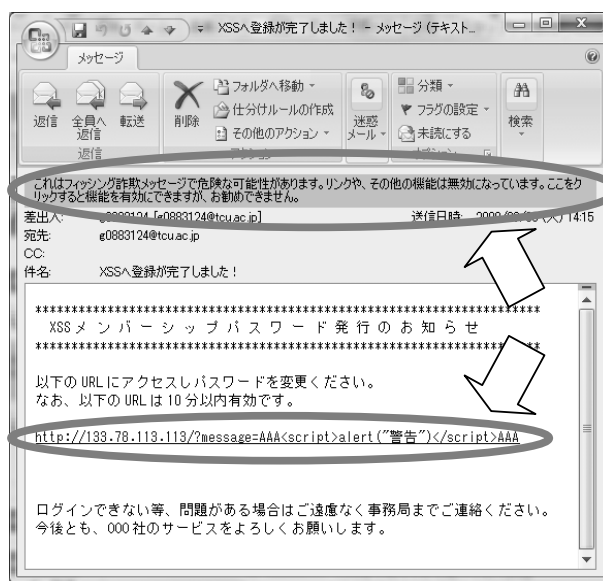


図3 従来のアドバイス例<sup>[3]</sup>(注1)

従来のメールソフトでは、情報支援としてフィッシングメールなどの危険なメールに対して警告を表示する機能がある。図3はフィッシングメールを受信した例である。差出人などのヘッダ情報を表示する部分に閲覧中のメールがフィッシングメールである可能性を指摘したアドバイスを表示している。また、本文中に含まれる危険なURLとプログラムを無効化することにより、安全性の確保もされている。

しかしながら、表示されるアドバイスは非常に短い説明文であるため、危険性に対する具体的な指摘がされていない。

### 3 提案する情報支援システム

#### 3.1 情報支援システムとは

本研究で提案する情報支援システムとは、文章解析技術と情報推薦技術を活用することによりメール閲覧画面を動的に配置し、ユーザのメール閲覧行動を支援するシステムである。具体的な動作内容としては、メールの受信時に件名や本文中に記載されているテキストの内容をPHP5でサポートされている正規表現でマッチング処理し、メールに記載されている内容を判別して内容ごとに分類する。

次に、分類された内容をもとにしてメール一覧表示画面を動的に構築する。そして、ユーザが選択した分類のメールを画面の任意の位置へ推薦することで、選択した分類のメールが他分類のメールにより隠されることを防止する。また、メールの本文閲覧時には、文章解析によりメール文中に含まれる脅威を判断して危険性の内容に合わせたアドバイスを表示する。

以上の情動的な支援により、ユーザのメール閲覧行動の円滑化とユーザ行動の補助を実現する。

#### 3.2 システム構成

本研究で提案するシステムを図4に示す。本研究ではWEBアプリケーションとして情報支援システム(WEB Mailサーバ)をApache2.2, PHP5, PostgreSQL8.2で作成した。なお、開発言語にはHTMLとPHP5を使用した。

本システムでメールを受信・表示するときメールの内容(テキスト)をPHPによる正規表現でマッチング処理して文章解析することにより、メールの内容に合わせて分類する。分類結果に基づいてメールを推薦することにより、重要分類以外の迷惑分類や私事分類のメールによって重要なメールが隠される可能性が低くなるようにシステムを開発した。

また、メール文章に応じてアドバイスを表示するアドバイス表示機能もあり、フィッシングやクロスサイトスクリプティングなどを含む危険なメールに対して詳細な

情報提供が可能になっている。

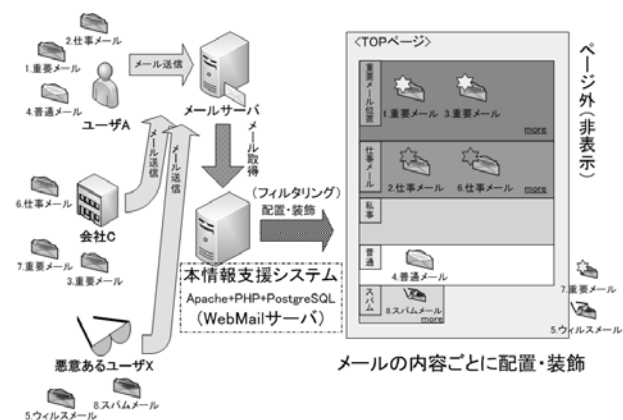


図4 提案システムの構成

#### 3.3 メールの推薦機能

メール推薦機能としては、図5に示すように分類結果ごとにメールをそれぞれ指定した場所へ配置するようにした。分類ごとの推薦件数は初期設定で“重要”7件、“仕事”5件、“私事”3件、“普通”3件、“迷惑”2件で、合計20件である。また、分類ごとに背景を装飾することで分類結果を一目で判別できるように設定した。

No.	件名	From	受信日	サイズ
1	重要メール	重要メール	2009年1月10日(土) 14:03:25	247Byte
2	仕事メール	仕事メール	2009年1月10日(土) 14:03:25	247Byte
3	私事メール	私事メール	2009年1月10日(土) 14:03:25	247Byte
4	普通メール	普通メール	2009年1月10日(土) 14:03:25	247Byte
5	迷惑メール	迷惑メール	2009年1月10日(土) 14:03:25	247Byte
6	重要メール	重要メール	2009年1月10日(土) 14:03:25	247Byte
7	仕事メール	仕事メール	2009年1月10日(土) 14:03:25	247Byte
8	私事メール	私事メール	2009年1月10日(土) 14:03:25	247Byte
9	普通メール	普通メール	2009年1月10日(土) 14:03:25	247Byte
10	迷惑メール	迷惑メール	2009年1月10日(土) 14:03:25	247Byte
11	重要メール	重要メール	2009年1月10日(土) 14:03:25	247Byte
12	仕事メール	仕事メール	2009年1月10日(土) 14:03:25	247Byte
13	私事メール	私事メール	2009年1月10日(土) 14:03:25	247Byte
14	普通メール	普通メール	2009年1月10日(土) 14:03:25	247Byte
15	迷惑メール	迷惑メール	2009年1月10日(土) 14:03:25	247Byte
16	重要メール	重要メール	2009年1月10日(土) 14:03:25	247Byte
17	仕事メール	仕事メール	2009年1月10日(土) 14:03:25	247Byte
18	私事メール	私事メール	2009年1月10日(土) 14:03:25	247Byte
19	普通メール	普通メール	2009年1月10日(土) 14:03:25	247Byte
20	迷惑メール	迷惑メール	2009年1月10日(土) 14:03:25	247Byte

図5 メールの推薦画面

メールの推薦により、同じ新着メールの中でも重要なメールがその他のメールにより隠される可能性が低くなっている。

#### 3.4 提案するアドバイス表示

アドバイス表示機能では、文章を解析することによりメールの内容を判断し、内容に合わせたアドバイスを表示した。また、文中に含まれるURLのハイパーリンクを自動的に構築せず、危険なプログラムの部分を強調表示してユーザの注意を促している(図6参照)。

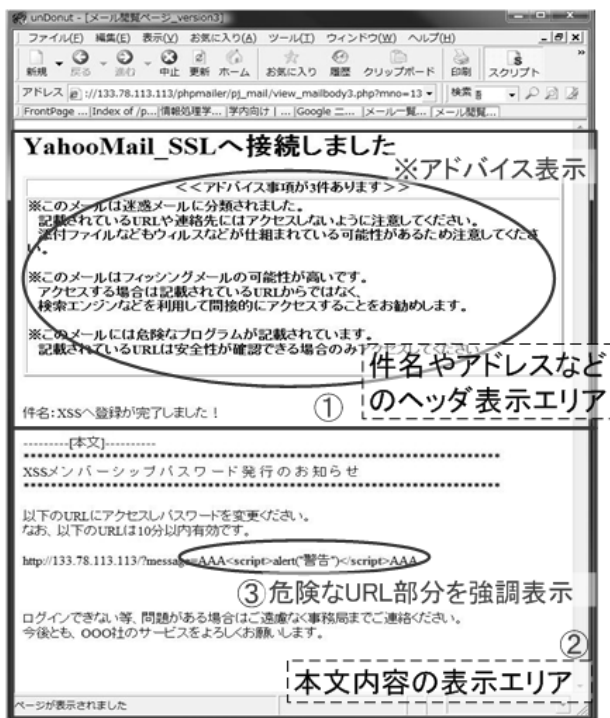


図6 提案システムでのアドバイス表示

## 4 実験1：表示機能のシミュレーション実験

### 4.1 実験1の概要

作成したシステムの特徴を明らかにするため、メールの表示状況のシミュレーション実験を行った。実験では、重要分類など、ユーザが表示設定した分類を想定した“分類I”と、それ以外の分類を想定した“分類I以外”の2種類のメールサンプルを用意し、分類Iの割合・分類精度・表示件数設定の3パラメータを変化させ、分類Iの表示件数の変化を計測した。

実験の計測結果を従来の20件新着リスト方式と比較をすることにより、どの程度の分類精度と表示件数設定であれば提案方式が従来方式より有効な表示結果となるかを明らかにするための実験を行った。

### 4.2 実験1の結果

サンプル数50件と100件の各実験結果を図7, 8に示す。実験結果によると、サンプル数が50件の場合は分類精度が50%以上なければ分類Iの表示件数が従来方式を上回ることができない結果となった。

また、サンプル数が100件の場合は50件の場合よりも提案手法が従来よりも上回る結果となった。さらに、表示件数設定した値に従って分類Iが表示される件数の上限が正しく設定されていることも確認することができた。

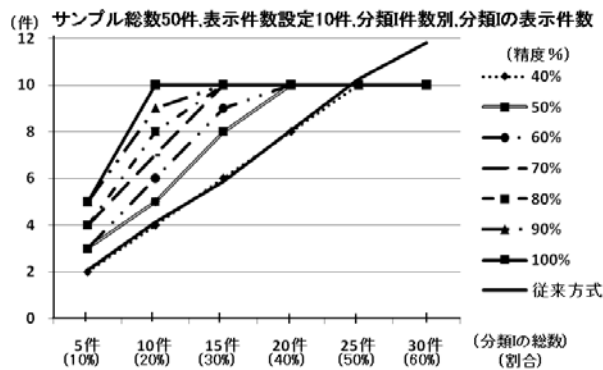


図7 サンプル数50件における表示件数

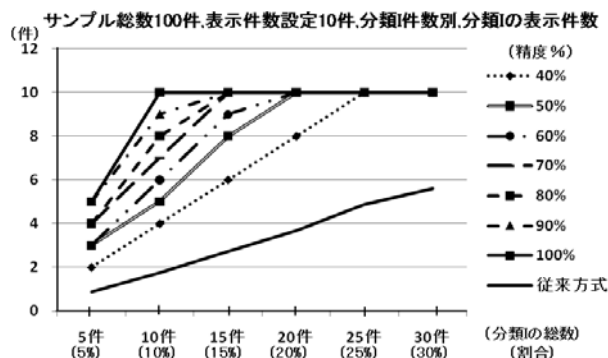


図8 サンプル数100件における表示件数

### 4.3 考察1

実験1の結果から、サンプル数が50件の場合は50%以上の分類精度があれば従来よりも表示件数を多く推薦できること、表示件数設定により表示上限を任意に制御可能であることが明らかになった。さらに、サンプル数が100件に増加すると提案手法が有利になるため、大量のメールを受信する環境で効果を発揮しやすいことが明らかになった。

また、分類Iの割合が全体の20%であれば分類精度50%以上で表示件数設定4件前後が従来よりも表示件数を大きく、もしくは少なくする条件であることなどが分かった。実験1により、推薦設定した分類の表示件数を従来方式より増減させるための条件が明らかになった。具体的な条件は、分類Iの含まれる件数(割合)別の分類精度と表示件数設定である。

## 5 実験2：メール送信実験

### 5.1 実験2の概要

実験1の結果から提案手法が従来よりも有利となる条件が判明した。実験1で明らかになった条件が妥当であるか、より実際の使用環境に近い条件でメールの送信実験を行った。

実験2では、分類を重要、仕事、私事、普通、迷惑の5種類設定し、各種類30件のサンプルを用意してサンプル総

数150件で実験した。送信するときは各分類のサンプルからランダムな件数抽出し、ランダムな順序で送信することにより、より実際の利用環境に近いと考えられる条件で実験した。

提案方式の画面表示条件は、従来方式の新着20件リスト表示と提案方式(重要7件, 仕事5件, 私事3件, 普通3件, 迷惑2件の合計20件表示)の2つで行った。

### 5.2 実験2の結果

抽出回数別における実験結果と平均での実験結果を表1, 図9に示す。提案方式(理想)とは分類精度が100%であった場合を想定したもので、提案方式(実測)は実際に表示されている種類と件数を測定したものである。

表1 抽出2回目の送信データ

分類	送信数	分類ミス	分類精度	全体の割合
重要	11	3	73%	17%
仕事	13	5	62%	20%
私事	17	3	82%	27%
普通	14	1	93%	22%
迷惑	9	1	89%	14%
合計	64	13	80%	

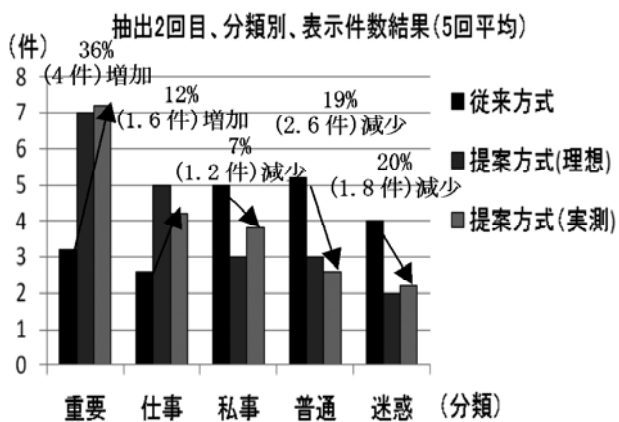


図9 抽出2回目の表示結果

表1, 図9は抽出2回目の実験結果である。分類のうち重要, 仕事で表示件数が増加し, 私事, 普通, 迷惑の各分類では表示件数が減少する結果となった。具体的な各分類が表示される割合(注2)の変化数値としては, 重要分類で36%(4件)増加, 仕事分類で12%(1.6件)増加となった。一方, 私事分類では表示が7%(1.2件)減少, 普通分類では19%(2.6件)減少, 迷惑分類では20%(1.8件)減少となった。

表2 抽出3回目の送信データ

分類	送信数	分類ミス	分類精度	全体の割合
重要	17	3	82%	21%
仕事	17	7	59%	21%
私事	12	4	67%	15%
普通	16	3	81%	20%
迷惑	19	2	89%	23%
合計	81	19	77%	

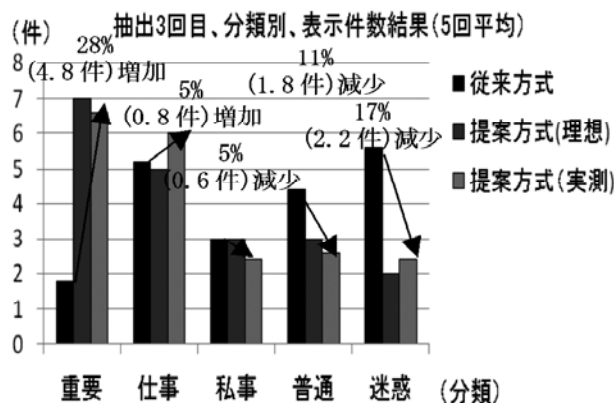


図10 抽出3回目の表示結果

表2, 図10は抽出3回目の実験結果である。この結果では, 重要分類で28%(4.8件)増加, 仕事分類で5%(0.8件)増加となった。一方, 私事分類では表示が5%(0.6件)減少, 普通分類では11%(1.8件)減少, 迷惑分類では17%(2.2件)減少となった。

表3 5回平均の送信データ

分類	送信数	分類ミス	分類精度	全体の割合
重要	11.8	2.8	76%	16%
仕事	15.8	4.2	73%	22%
私事	14.6	3.6	75%	20%
普通	15.8	3	81%	22%
迷惑	14.8	1.2	92%	20%
合計	72.8	14.8	80%	

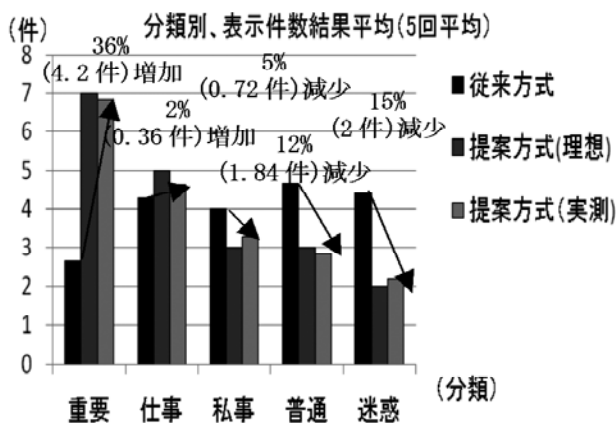


図11 表示結果(平均)

表3, 図11は実験結果を全て平均した結果である。平均した実験結果では、実験1の条件を受けて表示件数が増加するように設定した重要分類では、従来方式では重要メールの22%が表示されていたが、提案方式では58%が表示され表示件数が36%(4.2件)増加する結果となった。仕事分類でも2%(0.36件)表示件数が増加した。一方、表示件数設定が少なくなるように設定した私事分類では従来よりも表示件数が5%(0.72件)減少する結果となった。普通分類では表示件数が12%(1.84件)減少した。迷惑分類では従来方式では迷惑メールの30%が表示されていたのが、提案方式では15%表示される結果となり、15%(2件)少なく表示される結果となった。

### 5.3 考察2

重要メールは全体からみた割合が約15%であるため、実験1の結果を基準に考えると表示件数が3件以上であれば従来よりも多い表示結果となる予想と一致した。また、全体の約20%を占める迷惑メールに対しても同様であり、表示件数設定4件より少ない2件に設定した迷惑メールでは表示件数が減少する結果となった。そのため、実験1で明らかになった条件で表示件数設定行えば、ほぼその通りに表示件数を制御することが可能であることが明らかになった。

しかしながら、実験2では実測値が理想条件よりも前後する結果も得られ、予測通りにならない部分も判明した。この予想通りにならなかった部分は、分類ミスによる推薦漏れだけではなく、分類ミスされたメールが誤った種類へ分類されたため、誤推薦による表示がされたためである。しかし、誤推薦による表示件数の増減は1件以内であったため、実験結果全体への影響は少ないものと考えられる。

## 6 おわりに

### 6.1 まとめ

本論文では、ユーザのメール閲覧行動を情動的に支援することを目的にWEBメール受信システムに文章解析機能を組み込むことで、従来システムの問題点を解決した情報支援システムを作成し、実証を行った。

その結果、実験1で明らかにした条件(分類Iの割合が全体の20%であれば分類精度50%以上で表示件数設定4件前後が従来よりも表示件数を大きくする、もしくは少なくする条件など)に従ってメール表示件数設定することにより、従来方式よりも柔軟にメール表示画面を設定することが可能であることが明らかになった。

実際、実験2の平均結果では重要分類で表示件数を36%増加させる結果となった。迷惑分類では、表示件数を15%減少させる結果となった。以上のことから、表示件数の制御が実験1の条件が表示件数の制御に適用できるこ

とを確認できた。

### 6.2 今後の課題

今後は評価方法を検討し、評価手法の妥当性を調べる必要がある。特に、今回設定した分類区分は恣意的に決定したものであり、一般化が十分ではないと考えられる。従って、分類区分の定義や評価方法など今後も検討が必要である。

また、本システムを改良し、複数用意した分類をユーザが任意に選択・設計したり、表示件数設定もGUIなどを通して自由に設定することができるようにして汎用性を向上させる必要がある。

### 謝辞

本研究を進めるにあたり数多くの貴重なご意見を頂きました。東京都市大学環境情報学部情報メディア学科奥平雅士教授、ならびに、大谷紀子准教授には深く感謝の意を表します。そして、お世話になりました諏訪研究室の皆様へ感謝致します。

### 注

- (注1) アドバイス機能を搭載しているメールソフトはOutlook2007のみであったため、比較対象として本ソフトの画面を取り上げた。
- (注2) 各分類が表示される割合とは、表示対象分類の何%が画面に表示されているかを示した数値である。

### 参考文献

- [1] 総務省, H20年通信利用動向調査(世帯編), 2008.10
- [2] 河重貴洋, 大島裕明, 小山聡, 田島敬史, 田中克己, 情報処理学会 研究報告, コンテキストを用いたメールの情報補完, 2006.7  
[http://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR200800\\_001.pdf](http://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR200800_001.pdf)
- [3] Microsoft, Microsoft Office Outlook 2007,  
<http://office.microsoft.com/ja-jp/outlook/default.aspx?ofcresset=1>