

## 送受信履歴と情報抽出に基づく電子メールの個人適応型ランキング

長谷川 隆明

NTT サイバースペース研究所

〒 239-0847 神奈川県横須賀市光の丘 1-1

Tel: 0468-59-8214

hasegawa@isl.ntt.co.jp

### あ ら ま し

電子メールの処理に対するユーザの負荷を軽減するためには、メッセージにプライオリティを割り当てて、電子メールをランキングすることが有効である。本稿では、ユーザの個人性とメッセージの重要性を反映した電子メールのランキングの手法を提案する。本手法の特徴は、メッセージの送受信の履歴を用いたユーザの個人性の学習とメッセージの本文を対象とした情報抽出によるメッセージの属性の獲得と、重回帰分析により決定される重み係数を用いた各属性値の重み付き総和によるプライオリティの計算である。評価実験の結果、新規に受信した電子メールのランキングや重要なメッセージのフィルタリングにおいて高い精度が得られ、本手法の有効性が確認された。

### キーワード

個人適応, 情報抽出, 電子メール, プライオリティ, ランキング, フィルタリング,

## Personalized E-mail Ranking Based on Communication History and Information Extraction

Takaaki HASEGAWA

NTT Cyber Space Labs.

1-1 Hikarinooka, Yokosuka, Kanagawa 239-0847, Japan

Tel: +81 468 59 8214

hasegawa@isl.ntt.co.jp

### Abstract

A problem in e-mail communication is that users find it difficult to quickly discern the really important e-mail messages from among the many e-mail messages received. Accordingly, a method for assigning priority to each message received is proposed that automatically extracts the message's features from the message and forms the personal profile of each user. The priority of a message is calculated as the weighted sum of these features using the weights of each feature determined by multiple-regression analysis. An experiment demonstrates that the proposed method can be put to practical use such as ranking or filtering the many messages received.

### key words

personalization, information extraction, e-mail, priority, ranking, filtering

## 1 はじめに

インターネットの普及により、電子メールは今や現代社会に欠かせないコミュニケーションの手段となった。電子メールの利点は、一度に多くの人に送信ができる同報性や、相手とのコミュニケーションのための時間を選ばない非同期性にある。ところが、ユーザが受信する電子メールには、緊急に読まなければならないメッセージなど高いプライオリティを持つ電子メールから、ジャンクメールなどの低いプライオリティを持つ電子メールまで広範囲に渡る。このような状況においては、多数の電子メールを受信するユーザにとって、重要な電子メールを見落とす危険性が存在している。この危険性を回避するためには、電子メールにプライオリティを割り当てて、電子メールのランキングを行うことが必要となる。

本稿では、受信側のユーザの個人性とメッセージの重要性を考慮した電子メールのランキングについて、次の3つのステップからなる手法を提案する。最初のステップで、ユーザの電子メールの送受信の履歴に基づいてユーザの個人性を学習する。次のステップで、メッセージの本文を対象とした情報抽出を行い、メッセージ内の特定の表現から本文の重要性に関する特徴を抽出する。最後のステップで、学習により得られた個人性と情報抽出により得られた本文の重要性に関する特徴を用いて、新規に受信した電子メールにプライオリティを割り当てる。

本稿の構成は、以下の通りである。まず第2節で電子メールの処理に対するユーザの負荷を軽減するための従来のアプローチについて述べる。第3節で電子メールにプライオリティを割り当てる方法を提案し、第4節で実験結果について報告する。第5節で実験結果に対して考察し、最後に第6節で結論を述べる。

## 2 従来のアプローチ

電子メールの処理に対するユーザの負荷を軽減するためのアプローチとして、まず挙げられるのは、電子メールシステムと個人情報管理システムの融合である。電子メールを受信する頻度の高いユーザは、“inbox”などの一つのフォルダにメッセージを蓄積する傾向がある。これは、メッセージをカテゴリに分類して別のフォルダに格納する時間がないことや、同じような基準で継続してメッセージを分類することが難しいことに由来している。一旦蓄積されたメッセージのフォルダには、予定やToDoなどユーザの個人情報が含まれているが、蓄積されたメッセージが多すぎると個人情報を管理することが難しくなる。これを回避するために、メッセージにマーカーを付け

たり、リマインダを設定したりする機能を持つシステムが提案されている [Whittaker 1996]。しかしながら、このシステムでは、これらの機能を適用する重要なメッセージの検索はユーザに委ねられている。すなわち、高いプライオリティを持つメッセージの自動認識は行われていない。

一方、重要なメッセージのフィルタリングにおいては、送信側のユーザによって付与されたプライオリティを利用する方法が、インターネットで流通しているソフトウェアや市販のソフトウェアでは一般的になっている [Netscape 1997, Microsoft 1997]。この手法は以下の電子メールの特徴を利用している。電子メールは、封筒としての役割を持つヘッダとメッセージの内容としての役割を持つボディからなっている。ヘッダの詳細はRFC822[RFC822 1982]の規格により定義され、特定の意味を持つタグの集合から構成される。例えば、“To:”や“From:”, “Subject:”というフィールドは電子メールの宛先と差出人のメールアドレスと電子メールの主題をそれぞれ表している。このような定義されたタグ以外にも、ユーザが自由に定義することが許された“X-”で始まる拡張フィールドが存在する。プライオリティを表すための拡張ヘッダとして、通常“X-Priority:”が用いられている。これらのヘッダのフィールドやプライオリティを表す拡張フィールドを用いることによって、メッセージのフィルタリングが行われる。しかしながら、このフィルタリングの方法では、次のような問題点が存在している。まず、ヘッダの“Subject:”フィールドでは、キーワードを設定しなければならないが、ユーザにとって適切なキーワードを設定することは難しい。また、電子メールのボディに記述されているメッセージの本文を参照しないため、ヘッダの“Subject:”フィールドにメッセージの内容と掛け離れたキーワードが存在する場合や、本文に重要な記述が存在する場合には限界がある。さらに、拡張フィールドにおけるプライオリティは、送信側のユーザが設定したものであるため、受信側のユーザの個人性を無視したものとなっている。

サービスを受ける側のユーザの視点に立った個人適応の方法は、WWWを対象としたユーザの支援システムにおいて盛んに研究されている。そのようなシステムの典型は、WWWのサイトをユーザに推薦するシステム [Levy et al 1996, Lieberman 1995] である。これらのシステムは、ユーザにより入力されるキーワードを用いたり、ユーザがアクセスしたページの履歴を利用したりするアプローチを採用している。しかしながら、これらのアプローチは、WWWを対象とした場合に限って有効であるため、電子メールを対象とすることはできない。

電子メールにおける個人適応においては、ユーザがメッセージを処理するときの状況や行動を学習するアプローチが有効であるとして、このアプローチを採用した研究に、Mail Agent[Maes 1994]がある。Mail Agentは、メッセージの読まれる順序や削除などのユーザの行動を観察し、その行動を取ったときの状況を記憶することによって、Memory Based Reasoningによりユーザの電子メールに対する習慣を学習する。ここでの状況は、メッセージの差出人やサブジェクト、日付などのヘッダに存在する属性の集合からなっている。学習の過程では、過去に取ったユーザの行動とそのときの状況のすべての対が記憶され、過去のすべての状況と現在の状況との距離が各属性の値の差異の重み付き総和により計算される。現在の状況と過去の最も近い状況との距離の値が閾値を越えると、過去の最も近い状況で取った行動を現在の状況においてユーザが取るべき行動として、自動的に実行したりユーザに提案したりする。しかしながら、この手法の問題点は、ユーザへの動的な適応のために定期的に行われる重み係数の学習のための計算が、状況を構成する属性が多いため時間がかかることと、状況を構成する属性はすべてヘッダの情報を用いているため、メッセージの本文に記述されている情報を考慮に入れていないことである。

そこで、本稿では、ユーザの電子メールの送受信の履歴から得られるユーザの個人性と、メッセージの本文から得られるメッセージの重要性を反映した電子メールのランキングの手法を提案する。

### 3 電子メールのランキング

#### 3.1 プライオリティの定義

電子メールのランキングを行うためには、まず、どのようなメッセージが高いプライオリティを持つのかということ定義しなければならない。メッセージの持つプライオリティのレベルについて、我々は次のように考えている。プライオリティの高いメッセージとは、個々のユーザによって異なる個人性の違いによって、様々に異なったメッセージを指すであろう。例えば、あるユーザにとって密接な関係にある差出人からのメッセージや、そのユーザが興味を持っているトピックを含むメッセージなどは、そのユーザにとってプライオリティが高いと考えることができる。これに加えて、ユーザに対する期限付きの依頼や予定の通知などが本文に含まれているメッセージもプライオリティは高いと考えられる。例えば、メッセージに含まれている依頼（ユーザに対するToDo）の期日が近いときには、ユーザはすぐにメッセージを読まなければならないであろう。メッセージの持つこれ

らの属性もプライオリティの割り当てに取り入れなければならない。

電子メールのプライオリティの割り当てを行うに当たって、ユーザの個人性やメッセージの重要性を示す次の5つの属性を用いる。

- メッセージの差出人
- メッセージの対象者（ユーザ宛やメーリングリスト宛）
- メッセージのトピック
- メッセージのタイプ（予定の通知やユーザに対するToDo）
- 予定の期日やToDoの期限に基づく緊急性

これらの属性を考慮することにより、例えば、疎遠になっている知人からの挨拶のメッセージよりも、期限の近いToDoを含む上司からのメッセージが優先される。

#### 3.2 メッセージからの情報抽出

メッセージのトピックとメッセージのタイプ、緊急性を得るために、メッセージに含まれる予定やユーザに対するToDoを抽出しなければならない。そこで、予定の期日、場所、タイトル、ToDoの行動、期限等からなるテンプレートを用意することによって、パターンマッチングによる情報抽出を行う[長谷川 1998]。パターンマッチングに用いるパターンは、キーワードや文字の種類である。電子メールでは、誤字脱字や文法間違いなどが存在してテキストの質が低いので、パターンマッチングの際に形態素解析は行わない。図1にパターンの例を示す。展開されるパターンの数はおよそ1,000個である。これらのパターンマッチングに加えて、箇条書きの表現からも情報を抽出している。

抽出された予定のタイトルは、ヘッダのSubject: フィールドと同様にメッセージのトピックを表していると考えることができる。メッセージのトピックにおいては、おおよそ同じトピックを表していても、表記が異なっている場合には、表記の差異を吸収することが望ましい。例えば、“セミナー”、“講習会”、“Seminar”の各表記は同じトピックを表すと考えることができるので、これらは同じひとつのトピックとして扱うことが自然である。本稿では、予定のタイトル及びヘッダのSubject: フィールドに含まれる特定のキーワード群が表す概念を、メッセージのトピックと定義する。予定のタイトルに複数のトピック（例えば、AとB）が存在する場合は、メッセージのト

```

<Schedule> ::= {(<Start_rep>(<End_rep>)?(<Location_rep>)?
<Event_rep><Object_pp><Inform>
|(<Start_rep>(<Location+>)?“の”)?
<$Event+>(“の件”?{“について”|“です”})}
<Inform> ::= {“開催”{({“致”|“いた”})?{“します”|“したい”}}|“する”}
|“行”(“な”?“われ”{“ます”|“る”})}
<Event_rep> ::= (<Modify_rep>)?<$Event+>
<Modify_rep> ::= <任意の文字>+{“とした”|“についての”}
<ToDo> ::= {(<Deadline>)?(<Contact>)?<Request>|<Deadline>}
<Request> ::= ({“ご”|“御”})?{“回答”|“連絡”|“報告”|“提出”}(“を”)?
|{({“して”}?{“下さ”|“くださ”|“頂”|“いただ”})|{“願い”|“ねがい”}}
<Start_rep> ::= <Date_pat>({“から”|“~”|“より”|“スタート”|“-”})?
<End_rep> ::= <Date_pat>({“まで”|“に”|“にかけて”})?
<Location_rep> ::= <$Location+>{“におきます”|“において”|“にて”|“で”}
<Object_pp> ::= {“を”|“について”|“か”}
<Deadline> ::= {<Deadline1>|<Deadline2>|<Deadline3>}
<Deadline1> ::= {<Date_pat>{“まで”|“迄”|“中”}(“に”)?|<Date_pat>“を”
{“締”|“メ”}(“め”?{“切”|“き”}){“り”}?}
<Deadline2> ::= {“締め切り”|“メ切”|“期限”}{“は”|“か”|“を”}<Date_pat>{“まで”}
{“と”|“です”|“に”}
<Deadline3> ::= {({“大”}?“至急”|“今すぐ”)}
<Date_pat> ::= {({[0-9]+“年”|“平成”[0-9]+“年”])?[0-9]+“月”[0-9]+“日”
|[0-9]+“/”[0-9]+(“日”)?}({“<任意の文字>+”}<任意の文字>+
|({“午前”|“午後”})?[0-9]+“時”[0-9]+“分”|[0-9]+“.”[0-9]+“分”})?}

```

図 1: 情報抽出に用いるパターンの例

ピックを  $A$  と  $B$  および  $A \wedge B$  とする。また、トピックが存在しない場合は、“その他”というトピックを割り当てる。

メッセージのタイプは、情報抽出の結果、少なくとも一つの予定の通知やユーザに対する ToDo が含まれている場合に数値 1 を割り当て、そうでない場合には、数値 0 を割り当てる。ただし、予定の通知の有無とは、予定の日時と場所の両方が存在する場合に予定の通知と定義する。また ToDo の有無とは、質問と依頼を表す行動の表現のうち、“参照下さい”や“ご利用下さい”などの抽象的な表現以外が存在する場合にはすべて ToDo と定義する。

メッセージの持つ緊急性は、抽出された予定の期日または ToDo の期限の日時とメッセージの送信日時との差である残り時間を正規化し、表 1 に示すように 0 から 1 までの数値で表す。メッセージの送信日時として、ヘッダの Date: フィールドの値を用いている。複数の予定の期日または ToDo の期限が存在する場合は、最も残り時間が少ないものを採用する。

### 3.3 送受信履歴からの学習

ユーザと密接な関係にある差出人やユーザの関心のあるトピックなどのユーザの個人性を獲得するためには、ユーザに質問や設定を強要する手法ではな

表 1: 期限と送信日時との残り時間の正規化

残り時間	正規化される数値
3 時間以内	1.0
当日以内	0.9
1 日以内	0.8
2 日以内	0.7
3 日以内	0.6
7 日以内	0.5
14 日以内	0.4
21 日以内	0.3
28 日以内	0.2
28 日以上	0.1
期限無し	0

く、ユーザの行動を観察することにより学習する手法が求められる。そこで、我々は、ユーザがメッセージを送信する行動がユーザの個人性を表しているという仮説を立て、ユーザの電子メールの送受信の履歴に着目することによって、ユーザの個人性を学習する手法を提案する。この仮説の根拠として、ユーザと密接な関係にある人とは電子メールのやりとりが頻繁に起こり、ユーザが興味を持つトピックでは議論が盛んに行われる等の傾向が挙げられる。

表 2: メッセージのトピックに関する返信率 (例)

トピック	受信総数	返信数	返信率
会議	48	36	0.75
講演会	60	24	0.4
展示	25	8	0.32
ゴルフ	18	16	0.89
⋮	⋮	⋮	⋮
その他	32	4	0.125

提案手法の基本的な考え方は、各属性の値 (例えば、トピック A) に該当する受信したメッセージの総数を分母とし、このうちのユーザが返信したメッセージの数を分子とした割合 (これを返信率と定義する) をメッセージのプライオリティに反映させることである。さらに、ユーザが新規に送信するメッセージはユーザの個人性を表しているため、これに対しても、新規に受信したメッセージに対する返信と考えることによって、返信率を更新する。各属性の値について返信率を更新する際に、その値に該当するメッセージを受信したとき、返信率の表にタブルが存在する場合は受信総数をインクリメントし、そうでない場合は新しくタブルを設ける。ユーザが返信を行うときには、返信数をインクリメントする。

ユーザの個人性として、メッセージの差出人と対象者とトピックの3つの属性を考える。メッセージのトピックについては、情報抽出の過程によって得られるすべてのトピックに関して返信率を計算する。表 2 にトピックについての返信率の例を示す。トピックが複数存在する場合には、それぞれのトピックについての返信率を更新する。また、トピックが存在しない場合には、“その他”として返信率を更新する。差出人についても同様に、すべての差出人のメールアドレスに関して返信率を計算する。対象者についても同様に、すべての宛先のメールアドレスあるいはメールアドレスの集合に関して返信率を計算する。

### 3.4 プライオリティの割り当て

メッセージのプライオリティは、ユーザの電子メールの送受信の観察による学習とメッセージからの情報抽出によって決定される。図 2 に新規に受信したメッセージのプライオリティが割り当てられる過程を示す。

メッセージのプライオリティは、メッセージのトピックや差出人、対象者に関する返信率とメッセージのタイプ、メッセージの緊急性の各属性から決定される。ユーザの電子メールの送受信の履歴からの学

表 3: 重回帰分析による属性の重み係数

属性	重み係数
定数	0.2593
トピック	-0.0549
差出人	0.3477
対象者	0.2768
タイプ	0.1713
緊急性	0.4431

習により、抽出されたすべてのトピックに関して最も大きい値を持つトピックの返信率と差出人に関する返信率、対象者に関する返信率が得られる。また、メッセージの本文からの情報抽出により、予定の通知やユーザに対する ToDo の有無についてのメッセージのタイプと送信日時と抽出された予定の期日または ToDo の期限との差異から緊急性が得られる。新規に受信したメッセージのプライオリティ ( $P$ ) は、各属性の値 ( $V_i$ ) を用いて次式により計算される。

$$P = a + \sum_i w_i * V_i$$

ただし、 $a$  は定数で、 $w_i$  は重み係数である。ユーザの個人性を反映させるために、これらの値を可変にする。

### 3.5 属性の重み係数の決定

ユーザの個人性を反映されるために、ユーザがランク付けした電子メールの順序を正解とし、重回帰分析によって正解から重み係数を決定する方法を提案する。ユーザによりランク付けされたメッセージのプライオリティ ( $P'$ ) は、次式により 0 ~ 1 までの値に正規化される。

$$P' = 1 - \frac{R}{N}$$

ただし、 $N$  はメッセージの数で、 $R$  はその中のランク付けされた順位を表す。メッセージのプライオリティ ( $P'$ ) と各属性の値 ( $V_i$ ) が与えられれば、定数と属性の重み係数を決定することができる。

$$P' = a + \sum_i w_i * V_i$$

定数 ( $a$ ) と属性の重み係数 ( $w_i$ ) は、重回帰分析により決定される。

## 4 評価実験

提案した手法の有効性を確かめるために、評価実験を行った。実験では、著者の電子メールを用い、受

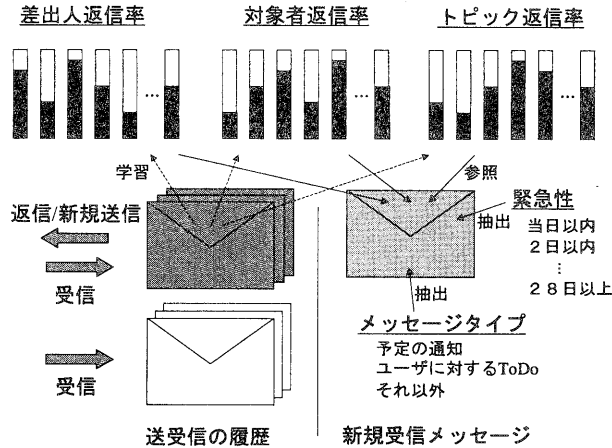


図 2: プライオリティの割り当ての過程

信した 2132 通と返信または新規に送信した 611 通を学習データとした。さらに、2132 通のうちの 250 通を正解としてランク付けを行い、属性の重み係数を決定した。そして、それ以降に受信した 236 通を評価データとして、ランキングおよび重要なメッセージのフィルタリングについての評価を行った。なお、実験に用いるトピックにおいては、149 種類を用意し、1 種類につき平均 5.8 個のキーワードを与えた。緊急性については、予定の日時と ToDo の期限を表 1 に従って正規化したものを用いた。また、返信率はメッセージが受信される時点での値を用いた。

学習データから、トピック、差出人、対象者の各属性に対する返信率の学習の後、学習データのうちの人手によりランク付けされた正解 250 通に対して、重回帰分析により決定された各属性の重み係数を表 3 に示す。得られた重み係数を用いたプライオリティの計算結果に従って、評価データを対象とした電子メールのランキングを行った。人手によりランク付けされた正解の順位と計算結果によるランキングの順位との差の頻度のヒストグラムを図 3 に示す。さらに、ランキングおよび重要なメッセージのフィルタリングについての結果を表 4 に示す。ランキングの評価については、人手による正解と計算結果との順位差の平均と標準偏差を用いた。また、フィルタリングの評価については、計算されたプライオリティの値が 0.6 以上の電子メールを優先すべき電子メールと考え、優先すべき電子メールのフィルタリングにおける再現率と適合率を学習データと評価データのそれ

それぞれについて求めた。

評価実験の結果、学習データおよび評価データのいずれも良好の結果を得ることができ、本手法の有効性が検証されたとと言える。

## 5 考察

以下では、評価実験の結果についての考察と今後の課題について述べる。

### 5.1 本手法の利点と限界

メッセージのランキングとフィルタリングに関して、学習データよりも評価データの方が精度に関してやや上回った結果が得られた。これは、学習データの中で人手による順位と計算結果による順位が大きく異なるケースがあっても、全体としてみれば、重回帰分析により決定された重み係数は、ユーザの個人性に関して一貫した傾向を表していると言える。このことは、ユーザ毎に一度だけ各属性の重み係数を決定しておけば、以降は同じ重み係数がそのまま使えることを示している。これにより、従来の手法では問題となっていた、重み係数の学習にかかる手間と時間が省略でき、本手法の利点が示された。

表 3 に示した、重回帰分析により決定された重み係数を参照すると、緊急性の重みが最も大きい。この種の情報は、通常メッセージの本文にしか現れない情報なので、メッセージの本文からの情報抽出が有効であることを示していると考えられる。また、差出人と対象者については、それぞれの返信率の学

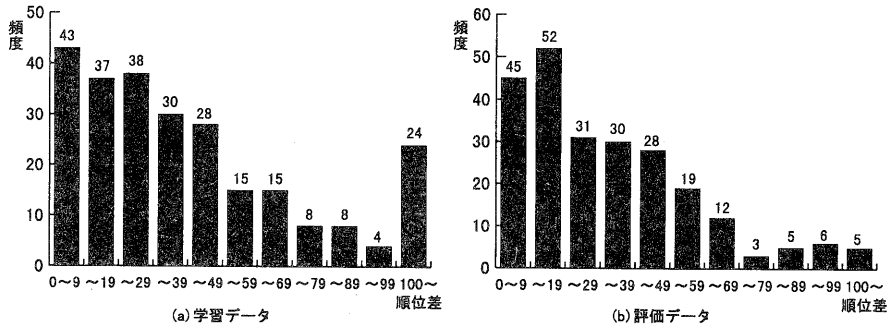


図 3: 順位差のヒストグラム

表 4: ランキングとフィルタリングについての評価

テストセット	ランキング		フィルタリング	
	平均	標準偏差	再現率 (正解数 / 全正解数)	適合率 (正解数 / 出力数)
学習データ (250 通)	41.2	34.9	67.0%(67/100)	78.8%(67/85)
評価データ (236 通)	33.1	29.1	80.9%(76/94)	76.8%(76/99)

習が効果的なため、適正な重み係数が決定されたと考えられる。メッセージのやりとりが多ければプライオリティが高く計算されるという点において、メッセージを送信するという行動がユーザの個性を表すという本手法の仮説が支持されたと言える。

逆に、トピックについては重み係数が小さく、ほとんどプライオリティの割り当てに影響を与えていないことがわかった。ここで、学習データと評価データに対する情報抽出の精度を表5に示す。評価データにおいては、タイプと緊急性の精度は非常に高いものの、トピックについては精度が下がっている。トピックの再現率については、トピックは予め設定するため、キーワードの漏れや種類の不足が精度を下げた原因として挙げられる。また、トピックの適合率については、文字列によるキーワードマッチングを行ったため、特にカタカナのキーワードにおいて誤って抽出される場合が目立ち、精度を下げた原因となった。トピックの抽出精度が上げれば、評価実験では小さかったトピックの重み係数が大きくなることも考えられる。

メッセージのタイプは、それほどプライオリティの割り当てには影響しなかった。これは、評価実験における情報抽出において、ToDoに相当する表現が存在すればそのままToDoとして抽出したことに依るためと考えている。例えば、“未提出の方は至急提出して下さい”等の表現は、ユーザの状況を判断す

ることができないため、ToDoとして取らざるを得ないという問題が存在している。

学習データにおけるフィルタリングの再現率が低下した原因は、ランキングにおいて順位差が大きいものが目立ったことにあると考えられる。これについては、評価データに比べて差出人や対象者についての頻度が少ないケースがあることや、学習データにおける情報抽出の誤り等が挙げられる。特に、表5に示したように、最も大きな重み係数を持つ緊急性についての情報抽出の再現率が低くなっていることが大きく影響していると考えられる。また、引用部分については、トピックと予定の通知の抽出の対象としたので、挨拶だけの様な内容に乏しいメッセージについても、引用された部分のトピックの返信率が反映されたため、人手による評価と大きく異なるケースも見られた。このようなケースは、ケースバイケースで重要であったりそうでなかったりするため、どう扱うべきか判断が難しい。

## 5.2 今後の課題

本手法は、人手によりランキングされたメッセージを必要とする。このため、ユーザが電子メールを使いながら簡単にメッセージをランク付けする環境が必要になり、そのためのインターフェースが用意されることが望ましい。また、このようなインターフェースは、ユーザの意図を反映した各属性の重み係数の再

表 5: 情報抽出の精度

属性	学習データ		評価データ	
	再現率	適合率	再現率	適合率
トピック	91.6%(837/914)	91.5%(837/915)	90.2%(754/836)	84.1%(754/897)
タイプ	96.4%(136/141)	98.6%(136/138)	99.3%(148/149)	100%(148/148)
緊急性	87.9%(58/66)	100%(58/58)	94.9%(74/78)	100%(74/74)

決定にも有効である。

また、個人性の学習のためのトピックは、予め設定しておかなければならないが、ユーザ毎に有効に機能するトピックの種類やキーワードを予め設定することは困難である。このため、トピックの情報抽出において、パターンマッチングと形態素解析による複合的なアプローチが必要であると考えている。

さらに、プライオリティの割り当てに用いる属性の過不足についても検討していきたい。各属性の重み係数は重回帰分析によって決定されるので、属性を増やすことは容易である。このため、どのような属性を追加すれば、プライオリティの割り当てがさらに有効になるのかを調べると同時に、そのような属性を抽出する情報抽出の手法についても検討していきたい。

## 6 おわりに

本稿では、ユーザの個人性とメッセージの重要性を反映した電子メールのランキングの手法を提案した。本手法の特徴は、メッセージの送受信の履歴を用いたユーザの個人性の学習と、メッセージの本文を対象とした情報抽出により得られるメッセージの重要性に基づくメッセージのプライオリティの計算である。個人性の学習では、メッセージのトピックと差出人、対象者のそれぞれについて、メッセージの返信の割合を獲得する。情報抽出では、個人性の学習に用いるメッセージのトピックに加えて、予定の通知やユーザに対する ToDo を表すメッセージのタイプと予定の期日や ToDo の期限を表す緊急性の程度を抽出する。メッセージのプライオリティは、履歴からの学習と情報抽出によって得られた各属性に対し、重回帰分析により決定された各属性の重み係数を用いることによって、これらの各属性の値の重み付き総和として計算される。評価実験の結果、一度ユーザ毎に決定した重み係数を用いたプライオリティの割り当てでは、新規に受信した電子メールのランキングや重要なメッセージのフィルタリングにおいて高い精度を得ることができ、本手法の有効性が確認された。

これにより、ユーザの電子メールは自動的にユーザの個人性やメッセージの重要性を反映してランキ

ングされ、ユーザの電子メールの処理に対する負荷は大きく軽減される。

## 参考文献

- [長谷川 1998] 長谷川隆明, 高木伸一郎: 電子メールコミュニケーションにおけるスケジュール情報抽出, 情報処理学会研究報告, NL123-10, pp. 73-80, 1998.
- [Levy et al 1996] Levy, A. L., Rajaraman, A., and Ordille, J. J.: Query-Answering Algorithms for Information Agents, In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pp.40-47, 1996.
- [Lieberman 1995] H. Lieberman: Letizia: An Agent That Assists Web Browsing, In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pp.924-929, 1995.
- [Microsoft 1997] Microsoft Corporation: Microsoft Outlook Express Reviewers Guide, 1997, <http://www.microsoft.com/ie/ie40/oe/oeexpress-f.htm>
- [Maes 1994] Maes, P.: Agents That Reduce Work and Information Overload, *Communication of the ACM*, Vol. 37, No. 7, pp. 30-40, 1994.
- [Netscape 1997] Netscape Communications Corporation: Guide To What's New: Netscape Communicator, 1997, <http://www.netscape.com/comprod/products/communicator/index.html>
- [RFC822 1982] Croker, D.: Standard for the Format of ARPA Internet Text Messages, STD11, RFC822, UDEL, 1982.
- [Whittaker 1996] Steve Whittaker and Candace Sidner: Email Overload: Exploring Personal Information Management of Email, *CHI96*, pp.276-283, 1996.