

電子メール利用履歴に基づいた処理順序取得システム

2 F - 2

加来田 裕和†

角 隆一†

NTT ソフトウェア研究所 ソフトウェア技術研究部
kakuda@slab.ntt.co.jp sumi@slab.ntt.co.jp

1 はじめに

近年、インターネット利用者の増加とともに、膨大な電子メールが送受信されるようになり、一日に 100~200 通受け取る人も珍しくなってきた。このため、「重要なメールがどれか分からない。すべて目を通す必要があり時間がかかる」といった、効率的な利用方法に関する問題が深刻化してきた。この問題に対して、電子メールを蓄積する場所を振り分けるツールや、重要性の判定要素を限定するなどのフィルタリング手法が提案されている。しかし、このようなフィルタ機能を用いても、候補を多数提示するなど不必要な情報が混入する問題や、特定の条件のドメイン等にしか適用できないという問題がある。

本稿では、電子メールの優先順位を推測し利用者に提示するフィルタリング手法の構築を目的として、利用者個々の電子メール利用履歴を基にした利用者プロフィールの表現方法について述べる。次に、そのプロフィールの属性である処理順序などを与えるシステムについて述べる。

2 優先度判定モデル

加来田[2]では、電子メールを選択する場合に必要な優先順位を判定する要素として、「期日」(「○月○日」や「×曜日」、「本日」など)、「緊迫した状態を表す表現」(「急いで」や「早急に」、「回答願います」など)、「情報発信者」(人名やメールアドレスなど、受信者との関係が分かる言葉)、「情報の内容を表す単語」(「進捗」や「NTT」など)、「情報到着日」(情報が到着した日付、または受信者が受け取った日付)を提示している。また、これらの要素を基にしてメール処理の緊迫の度合いと興味の度合いを算出し、電子メールの優先度を算出す

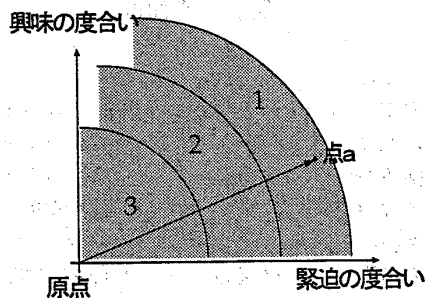


図 1 優先度判定モデル

る図1のモデルを提案している[1][2]。

しかし、優先順位判定要素の選定や値付けは、特定条件のドメインで効率的に機能する。そのため、他のドメインに適用する場合には要素や値付けを再設定する必要がある。

これを克服するために、利用者の操作を自動で取得し、利用者プロフィールを生成するシステムを作成し、そのプロフィールを基に優先順位判定要素を選択する手法を提案する。

3 利用者プロフィール

まず、利用者のプロフィールを表現するために以下のモデルを考える。

適当な期間をおいてメールを一括して処理する場合を前提に考えると、メールには処理順序があり、現在のシステム上以下の属性が取得可能である。

1. 「到着日時」-「参照日時」(Rt)

メールの読者は通常、到着しているメールのヘッダー見て、本文を読むかどうか判断する。この時、ヘッダーを表示した時を「到着日時」とし、メールの本文を参照したときを「参照日時」とした。メールの処理順序がこの間隔によって分かる。

2. 「相対参照順序」(Rr)

1回の処理中に参照した順序 r をヘッダー表示メール数 N で割った次の式で表される相対参照順序。(1を始点とする)

$$Rr = 1 - \frac{r}{N} \quad 0 < r \leq N$$

3. 「利用者指定順序」(Ar)

利用者が考えるそのメールの処理順序。具体的には利用者がメーラの機能を用いて指定する。

これらの値と、そのメールに出現する単語とその出現頻度によって、利用者個人のプロフィール P を以下のように表す。

$$P = \{P_a, P_b, P_c\}$$

$$P_a = \{(Rt, W, F_a)\}$$

$$P_b = \{(Rr, W, F_b)\}$$

$$P_c = \{(Ar, W, F_c)\}$$

W: 単語情報(単語 w の情報)

F_a : Rtにおける単語 w 出現頻度

F_b : Rrにおける単語 w 出現頻度

F_c : Arにおける単語 w 出現頻度

† Operation Sequence Capturing System based on Mail Server Log
Hirokazu KAKUDA Ryuichi SUMI
NTT Software Laboratories

ここで、 P_a, P_b, P_c は、それぞれ R_t, R_r, A_r を視点とした時の単語 w とその単語の出現頻度 F_a, F_b, F_c の集合を表す。

これらの順序 R_t, R_r, A_r において、それぞれの単語 w の出現頻度は異なると考えられる。順序 R_t では、全メールを通して絶対順序が取得できる。しかし、利用者が席を離れている場合等の要因を取り除くことが難しい。順序 R_r は、相対的にメールの順序を取得するので、順序 R_t と合わせて上記の要因を考慮可能である。また、順序 A_r は、利用者による絶対的な順序であるが、順序付けがなされない場合がある。そのため、順序 R_t, R_r を基に順序を考慮することが可能である。このように、それぞれの順序に対応するプロファイル P_a, P_b, P_c を取ることによって、利用者の行動をより多面的に取得できる。

ここで、単語情報 W の出現頻度 F は、単語情報 W が順序 R の複数の情報に出現する回数を表し、以下のように定める。

$$F = f(R, W)$$

R : それぞれ P_a, P_b, P_c のとき R_t, R_r, A_r
 W : 単語情報

また単語情報 W は、その単語そのもの w とその属性として品詞 C 、他の分類設定用の属性 U によって以下のように表す。

$$W = (w, C, U)$$

品詞 C は、名詞や動詞、その細分類とし、分類用の属性 U は、ある複数の単語に対する独自の頻度情報を収集できるような識別子とする。例えば、優先度判定モデルで扱った、「至急」「急ぎ」などの単語は属性「緊急」などとすることで優先度判定モデルと整合させることも可能となる。

4 処理順序取得システム

上記の利用者プロファイルにおける、電子メールの「到着日時」「参照日時」「相対参照順序」「利用者指定順序」を取得するため、処理順序取得システムを試作した。(図2)

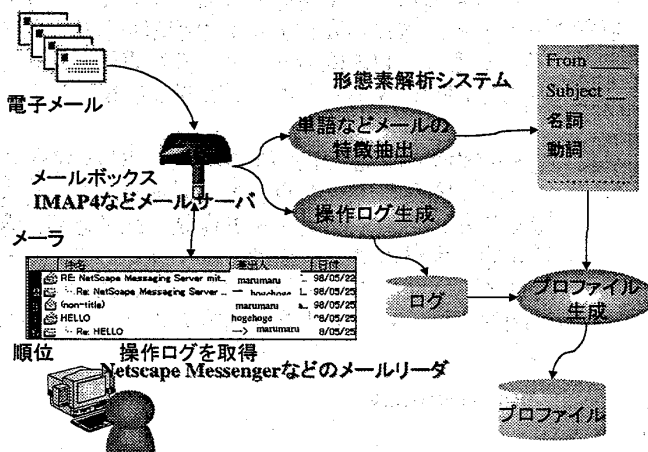


図2 処理順序取得システム

メールサーバより、発信者やサブジェクトなどのヘッダー情報を利用者が取得した時、そのヘッダーを送信したメールの数とそれぞれのメール ID、送信時刻を記録する。利用者が取得したヘッダーよりメールを選択してメールの本文を取得した時、そのメール ID と送信時刻を記録する。また、利用者が指定する順序は、取得したメールすべてに対して細かく順序を付けることは困難であること、既存のメールリーダーを使用することなどを念頭におき、5段階とした。また、メールを形態素解析システムにより、単語に分割する。

この後、順序 R のメールに出現する単語を、プロファイル P における順位 R の単語頻度情報に加算する。

以上のようにして、処理順序取得システムにより利用者のプロファイルを構成する。

5 本方式の効果

本方式の効果は以下の通りである。

1. プロファイルの精確化

a. 利用者個人に適応したプロファイルを構築可能。
 各個人の電子メール操作における3つの順序の属性を視点として、単語と、単語の出現頻度のみを基にしてプロファイルを構築するためである。

b. ドメインに依存しない自動要素抽出が可能。
 特定表現をあらかじめ定めず、利用者プロファイルの頻度情報に基づいて抽出を行うためである。

c. 日常的な操作のなかで利用者に負荷をかけることなくプロファイルを構築可能。
 日常的に行われる電子メール受信動作をモニターするためである。

2. 実装の容易性

a. 既存システムの組み合わせで構築可能。
 既存のシステムで収集可能なログとメール本文のみを用いているためである。

6 おわりに

本稿では、重要な電子メールを優先的に取得するフィルタリング手法の構築を目的として、電子メールの優先順位を判定する要素を自動的に選択するための、(1)利用者プロファイルの表現を提案した。また、(2)そのプロファイルの属性である処理順序などを取得するシステムについて述べ、このシステムの効果について考察した。

今後は、ここで仮定した利用者プロファイルの有効性の検証を行っていく。また、このシステムの性能について評価を行っていく。

参考文献

[1] 加来田: 時間的表現に基づく優先度判定モデル, 第55回情処全国大会, 1997
 [2] 加来田: 情報取得のための優先順位判定要素と順位判定モデル, 知能と複雑系, 113(13), pp.91-98 (Jul. 1998)
 [3] 森田: 情報フィルタリングシステム, 情報処理, 37(8), pp.751-758 (Aug. 1996)