

# メールヘッダー情報を利用した同報先推薦及び、誤送信防止方式の提案

境 美樹 新井 克也

日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

## 概要

企業におけるメールの誤送信は、個人情報や機密情報流出など大きな問題を引き起こす。これらは、アドレスの入力・選択ミスや送信先アドレスの確認不足など、人為的によるものが多い。そこで、仮名漢字変換で広く用いられている予測変換技術を応用して、ユーザの入力した送信先アドレスから同報相手の推薦、送信先アドレスの組合せの正当性検証を行う手法を提案する。

### 1. はじめに

インターネットの普及に伴い、企業における電子メールの利用率は100%近くに達している。このような電子メールの普及を背景として、「メール送受信によるウイルス感染」や「操作ミスによる誤送信」、「メールの不達」などの電子メール利用に際するトラブルも増加してきている[1]。

特に、企業におけるメール誤送信は、個人情報流出の原因や機密情報漏洩につながることから、その防止の手立てが必要とされている。

誤送信の主な原因は、アドレスの入力ミスなどヒューマンエラーによるものであると予想される。このようなアドレスの入力ミスは、アドレス帳のリストから選択すれば、ある程度、防止することができる。しかし、アドレス帳の登録数が膨大にあり、例えば、山田美香(yamada.mika@example.co.jp)と山田美貴(yamada.miki@example.co.jp)といったように、類似した氏名やメールアドレスの割合が高くなると、送信先を誤って選択する可能性が相対的に高くなる。このように、送信先が多岐にわたるような環境においては、アドレス帳使用による入力ミスの防止策だけでなく、送信先の誤選択を防止することが重要となる。

本研究では、特に送信数や送信先の数が大きくなりやすい企業におけるメール、特に複数人に送信する同報メールに着目する。これは、多人数に同報する場合、選択ミスが起こりやすいと考えるからである。そして、仮名漢字変換の一つである予測変換技術の「ある語彙の次にどの語彙が来る可能性が高いか?」という課題と解決手法を、メール作成時の「あるメールアドレスを選択した場合、同報する可能性が高いアドレスは何か?」という課題に応用することで、状況に応じた適切な送信先を推薦し、送信先の誤選択防止を試みる。

以降、2節で予測変換方式の概要を述べ、3節で同報先推薦及び誤送信防止方式について述べる。4節では、実データを用いた推薦方式の有効性を検証する。

### 2. 予測変換技術

#### 2.1 日本語入力の予測変換技術

一般的な日本語文字入力の場合、読みに該当する文字列を与えて仮名漢字へ変換する。しかし、携帯端末など入力インターフェイスに制限がある場合には、正確な文字列を入力することが困難である。そこで、少ない入力文字数で日本語入力を可能とする、予測変換技術が用いられている。

予測変換の挙動はインクリメンタルサーチ<sup>1</sup>と同様で、ユーザが文字を入力するたびにその先に来るだろう文字列を予測し、これを候補として列挙する。ユーザは、この候補の中から目的に合った文字列を選択すればよい。携帯電話やPDAなどに多数採用されているものとして、POBox[2]やT9[3]などがある。

#### 2.2 予測変換技術の応用上の課題

一般に、日本語入力における予測変換は、例文を解析した結果を予測辞書とし用意しておき、この辞書を使って、ある入力文字列に対して、それ以降に入力される確率の高い文字列や単語を予測し、これを優先して提示することで実現される。また、予測辞書から適切な候補が発見できなかった場合は、使用頻度の高い語や最近選択した単語の中から予測を試みる工夫もされている。

この技術をメールの送信先アドレスの推薦へ応用する場合、「例文」としてユーザが送信したメールのTo、

<sup>1</sup>検索したい単語を全て入力した上で検索するのではなく、入力のたび毎にリアルタイムに候補を表示させる検索手法。

Cc 及び, Bcc に記述された「アドレスの組合せ」を利用する方法が考えられる。

しかし, To, Cc, Bcc の各フィールドに書かれたアドレスの順序は, メール配信には影響が無いので, 一般の日本語文章のように単語順序に一般的な規則性や意味があるとは考え難い。例えば, Cc フィールドに 3 名のアドレスが入力されていた場合, その順序によって送信者の意図が異なると考えるより, その順序によらず 3 名に同報することを意図している, と考える方が自然である。以上の考察から, 日本語入力の予測変換技術をメールの送信先アドレスの推薦へ応用する場合, アドレスの順序を候補選択のパラメータとして使うことはできず, 推薦すべき送信先アドレス候補が多数になることが予想される。そこで, 本研究では, 入力されたアドレスに対する同報先の推薦に加えて, 推薦すべき候補が多数存在した場合の推薦候補の絞り込み手法も提案する。

### 3. 同報推薦及び, 誤送信防止方式

#### 3.1 利用データ

ユーザが過去に送信したメールのヘッダから以下の情報を抽出し, 先に述べた, 日本語入力の予測変換技術における「例文」として利用する。

- To に記述されたアドレス
- Cc に記述されたアドレス
- Bcc に記述されたアドレス

これらのデータは, 表1に示す通り, 送信したメールごとに記録しておく。なお, 表中のアルファベット a~e は, それぞれ独立なメールアドレスである。

表1 ヘッダ情報の例

	To	Cc	Bcc
メール1	a	b, c	
メール2	b	c, d	a
メール3	a, e	B	

#### 3.2 同報先推薦と送信先の妥当性検証

メール送信時, 送信先アドレスを確認するタイミングとしては, 次の2つが考えられる。

- (1) 送信先アドレスの入力時
- (2) メール送信直前

本稿で提案する手法では, ユーザの過去の送信履歴からアドレスの組合せを抽出し, (1)では, 送信先として妥当性の高いアドレスの組合せを推薦し, (2)では,

これから送信するアドレスの組合せの妥当性をチェックすることで誤送信の防止を行う。

以降, それぞれの詳細を述べる。

#### (1) 送信アドレス入力時

ユーザが送信先を選択し, 送信アドレスが 1 つ決まった時点で, 同報する可能性の高いアドレスの組合せを推薦する。候補が複数ある場合は, 同報頻度の高い組合せを上位に提示する。

推薦する同報先の抽出方法は以下の通りである。

##### 【推薦する同報先の抽出方法】

$A = \{a_1, \dots, a_n\}$ : 過去に送信したアドレスの集合

$S = \{E_1, \dots, E_m\}$ : 過去に送信した組合せの集合

ここで,  $To_i, Cc_i, Bcc_i$  をそれぞれ, To, Cc, Bcc フィールドに入力したアドレスの組合せだとしたとき,

$$E_i = \{To_i, Cc_i, Bcc_i, \text{送信回数}\}$$

であるとする。また,  $To(E_i) = To_i, Cc(E_i) = Cc_i, Bcc(E_i) = Bcc_i$  とする。

このとき, 1 つ目のアドレスとして

$$a_k (k \in \{1, \dots, n\})$$

を指定したならば,

$$S_1 = \{E_i \mid a_k \in To(E_i) \text{ and } E_i \in S\}$$

となる送信先の組合せ集合から, To, Cc, Bcc を宛先・同報先候補として推薦する。

例えば, 表1の送信履歴を利用した場合,

$$E = \{a, b, c, d, e\}$$

$$S = \{E_1, E_2, E_3\}$$

$$E_1 = \{\{a\}, \{b, c\}, \{\phi\}, 1\}$$

$$E_2 = \{\{b\}, \{c, d\}, \{a\}, 1\}$$

$$E_3 = \{\{a, e\}, \{b\}, \{\phi\}, 1\}$$

となり,

「a」を1つめのアドレスとして選択すると,

$$S_1 = \{E_1, E_2\}$$

となり, To の候補として「a, e」, Cc の候補として「b, c」「b」を推薦する。更に, Toとして「a, e」を選択すれば, Cc の候補も絞り込まれて, 「b」のみが推薦される。

このように, 1 つアドレスを決めれば, 同報頻度の高いアドレスを推薦することが可能となり, ユーザの入力の手間が省けるので, アドレスの誤入力/誤選択が防げる。

#### (2) メール送信直前

次にユーザが「メール送信」を実行した際, 送信先アドレスの組合せの妥当性をチェックする。

##### 【送信先の妥当性の検証方法】

送信しようとしているアドレスの組合せを

$$E_{\text{send}} = \{To_{\text{send}}, Cc_{\text{send}}, Bcc_{\text{send}}\}$$

とした時,

$$To_{send} \subset To(E_i)$$

$$Cc_{send} \subset Cc(E_i)$$

$$Bcc_{send} \subset Bcc(E_i)$$

を全て満たす,  $E_i \in S$  が存在すれば, 送信し, 存在しなければユーザに警告をする. つまり, 今まで一度も送ったことの無いアドレスの組合せで送信しようとした場合, 誤送信の可能性があるとして, ユーザに送信先アドレスの再確認を促し, 誤送信を防止する.

また, 警告したにもかかわらず, ユーザの判断で送信した場合は, その組合せをフィードバックし, 次回以降のメール作成のアドレス選択時には反映し, その組合せも正しいものとして扱う.

### 3.3 候補が多数ある場合

送信頻度の高いアドレスは, 多様な組合せで送信していると考えられる. すなわち, 送信頻度の高いアドレスが最初に指定されると, 推薦する候補が絞込まれず, リストに提示される候補数が多量になり, 誤選択を招く可能性がある.

そこで, 推薦する候補が多数ある場合には, 候補を提示する前に, ユーザが「キーとなるアドレス」を同報するか否かを選択することで候補の絞込みを行う.

ここで「キーとなるアドレス」とは以下の 2 つの基準に則ったアドレスを指す.

#### 【基準1: キーとなるアドレス】

$$S_i = \{E_h \mid a_k \in To(E_h) \text{ and } E_h \in S_{i-1}\}$$

を,  $i$  個めのアドレスを選択した結果, 推薦する組合せの集合とする, このとき,  $|S_i|$  をその要素数とする.

このとき, キーとなるアドレスとは,

$$|S_i| \text{ が最大になるような } a_k$$

を指す.

#### 【基準2: キーとなるアドレスの組合せ】

$A'$  をアドレスの集合  $A$  の部分集合とした場合,

$$S'_i = \{E_h \mid A' \in To(E_h) \text{ and } E_h \in S'_{i-1}\}$$

は,  $A'$  を選択した際に提示される候補の集合である.

このとき, キーとなるアドレスの組合せとは,

$$|S'_i| \text{ が最大になるような } A'$$

を指す. もし,  $To$  だけで絞込み切れない場合は,  $Cc$ ,  $Bcc$  にまで範囲を広げて絞込みを実施する.

例えば, 「a」に送信するときは, 同時に送信している可能性の高い「b, c」という組合せを提示する.

基準1・2で提示される候補を採択・棄却することで, 推薦すべき候補数の大幅な減少が期待できる. この様に, 絞込み結果を大きく左右するアドレスやアドレスの組合せを提示していくことで, 効率よく絞込みを行う.

一方, 基準 2 でアドレスの組合せが膨大にある場合は, 総当りで最大になるものを探すのは非効率的である. そこで, 本稿では次の手法を取る事にする.

#### 【方法1: アドレスの組合せの決定方法】

入力:  $a_1$  (メールアドレス)

出力:  $A$  (アドレスの組合せ)

1.  $A = \{a_1\}$
2.  $a_k = (A \text{ に対して基準1を満たすアドレス})$
3.  $a_{k2} = (A \text{ に対して基準1において2番目に } |S_j| \text{ が大きくなるアドレス})$
4.  $a_k' = (A \cup \{a_k\} \text{ に対して基準1を満たすアドレス})$
5.  $a_{k2}' = (A \cup \{a_{k2}\} \text{ に対して基準1を満たすアドレス})$
6.  $a_k = a_{k2}'$  かつ  $a_{k2} = a_k'$  ならば, 以下の処理を行う  
(ア)  $A = A \cup \{a_k, a_{k2}\}$   
(イ) 2へ戻る
7.  $A = A \cup \{a_k\}$
8. 終了

つまり, 基準1において,  $|S_j|$  の値上位 2 つを調べ, そのアドレスを選択された場合, その次に選択されやすいアドレスを予め調べておくことで, 組合せを作っていく. ユーザは, 提示された組合せ中のアドレスを全て選ぶ必要は無く, 同報したいアドレスのみを選択していけば, その先の候補は絞り込まれる.

これにより, ユーザがアドレスを入力/選択する回数は減少し, 提示される候補数も絞られるため, 誤入力/誤選択の可能性を低く抑えることができる.

### 4. 評価実験

本手法の有効性を検証するために, 4 名のメール送信履歴を用い評価を行った.

実験データとして, 各被験者から1年程度の送信履歴を収集しており, その送信メール数は表 2 の通りである.

表 2 被験者の送信メール数

被験者	送信メール数	送信先数	送信数 5 通以上のアドレス
A	6933 通	679	220
B	5522 通	522	219
C	2518 通	309	111
D	1043 通	193	64

これらデータのうち、送信数 4 通以下のアドレスを送信頻度が低いとすると、各被験者ともに約 65%は送信頻度が低い相手となる。つまり、これらのアドレスを選ぶと同報先候補はそれぞれ最大4つまでしか現れないので、誤選択の可能性は低い。

逆に、送信頻度が非常に高い相手では、同報先候補の数が 100 を超える場合もある。誤送信防止の観点で問題になるのはこのようなアドレスを選択した場合であると考えられる。

各被験者についての送信頻度の高い相手を選んだ際の平均候補数、及び最大候補数を表 3 に示す。今回、簡単のため、To フィールドに記載されたアドレスのみを対象とした。

被験者 A と B では、もし最大候補数になるアドレスを最初に選んでしまうと、正しい選択ができない程多数の候補が現れることになる。また、今回の実験データの場合、最大候補数になるアドレスは、送信頻度が高い傾向にあるため、被験者 A 及び B がこのようなアドレスを最初に選択する可能性は非常に高い。

表 3 5 通以上送信したアドレス

被験者	平均候補数	最大候補数
A	8.30	69
B	18.47	158
C	1.43	8
D	2.73	11

従来のメーラでは、複数アドレスに同報する場合、アドレス帳から一つずつ選ぶような形式だった。もし、アドレス帳に大量の登録アドレスがあるならば、そこから必要な分だけアドレスを選ぶ動作では、誤選択の可能性が高くなる。

しかし、3.2(1)の手法を利用すると、最初に正しいアドレスを1つ指定するだけで、同報先候補の絞り込みが行われ、最大でも表3の最大候補数まで減少することが分かる。アドレス選択時の選択肢の数を各被験者が元々持つ送信先数だとすれば、1/3 以下に減らすことが可能になるのである。

次に、被験者Bのように、同報先候補数が多数になるアドレスを持つ場合に着目する。

この時、送信回数の多い候補や、その候補を選んだ後の候補数が多いものを選んでいくと、選択しやすい候補数まで絞り込むためには、6~7回のアドレス選択が必要になる。

図1では、被験者Bが 1 つめのアドレスとして、最大候補数を持つアドレスを選択した上で、3.3 の基準1で

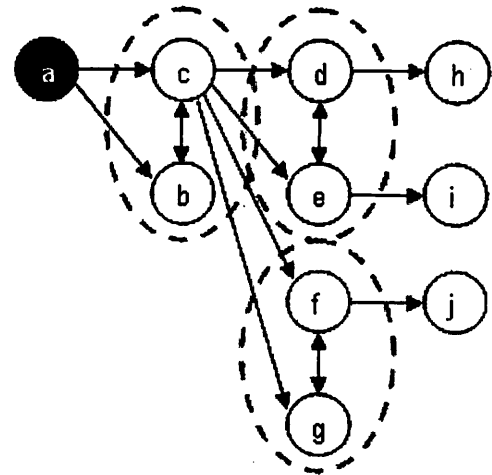


図 1 被験者Bのアドレス選択遷移

提示した手法を実施した場合の遷移の一部をグラフ形式で示したものである。

この図では、宛先として、「a」を選ぶと次に「b」又は「c」を選ぶ可能性が高く、「a, b」の次は「c」、「a, c」の次は「b」を選ぶ可能性が高いことを示している。つまり、「b, c」は、複数のアドレスを1つの組合せとしてみなせるパターンとなっており、3.3 の方法1で示した組合せの選び方が、基準2のアドレスの組合せに相当することが分かる。

ここで「a, b, c」と確定すると、同様に「d, e」「f, g」という組合せを候補としてあげることが可能である。

つまり、3.3 の基準2の組合せを利用した候補提示を行えば、今まで「a → b → c → d → e」という4回のアドレス選択が、2回に減少できるのである。

更に手法1を適用した場合の候補数の遷移を見るため、表 3 で示した最大候補数が特に多い、被験者 A と B に関して、それぞれ送信回数上位 5 位のアドレスを入力として、手法1を適用した結果を表 4 及び 5 に示す。

表 4 及び 5 のアドレスでは送信回数の順位、対応候補数はその順位のアドレスに対する候補の数、手法1適用要素数はその順位のアドレスに対して手法1を適用した結果候補として提示された組合せの要素数、全選択時の候補数は候補として提示された組合せの要素全てを選択した際の候補数を示している。

被験者 A の場合、元々の候補数が少ないため、手法1を適用しても、最大の効果が得られているとは言えない。一方、被験者 B では、元々持つ候補数が多いため、組合せとして提示される候補の要素数も多く、手法1の効果が得られている。また、今回の場合、提示される組合せの要素数も選択時の許容範囲内に収まっ

表4 方法1適用結果(被験者 A)

アドレス	送信回数	対応候補数	手法1適用結果要素数	全選択時の候補数
1	1192	51	2	5
2	1131	69	1	22
3	862	49	1	6
4	657	33	2	5
5	538	19	4	1

ており、提示された組合せの中から誤選択する可能性は少ないだろう。

手法1の効果がある／なしに関わらず、次に選択しそうなアドレスを提示することで、候補の数を減らせることは分かり、提案する手法が有効であると言える。

このように、メールのヘッダ情報を用いて送信先アドレスの絞込みを行うと、アドレスの選択回数を減少するので、選択ミスを防ぐことが可能となり、誤送信防止につながる。

## 5. 関連研究

電子メールを扱う研究には、メールのヘッダや本文を解析し、メール内に含まれる単語の重要度や単語間の相関ルールを導出する研究が行われており、多くのメールを解析すれば、個人のごとの特徴などが分かること報告されている[4]。これに比べて、本手法では、メール本文の解析は行っていないため、プライバシー侵害の問題は起こりにくい。また、処理自体も容易なので携帯端末など比較的処理能力が低い環境でも利用しやすいと考える。更に、機密情報保持を優先すべきビジネス用途のメールの場合は、本手法に付け加えて、送信メール本文の特徴と送信先アドレスの関係付けを行うことで、誤送信防止効果が高まると考えられる。

機密情報保持という観点での誤送信防止に対しては、メーラやサーバに、特定のキーワードを予め設定しておくことにより、そのキーワードにマッチするメールが送信されたときに警告又は送信中止を行う製品がある。同様に社外や登録外のアドレスへメールを送信しようとする場合に警告する製品もある[5][6]。これらは、予期しない相手に送信しないので有用だが、事前の設定や新たに専用サーバを置く必要がある。

## 6. おわりに

本稿では、メール誤送信防止を目的とした同報先推薦及び、送信先チェック手法の提案を行った。

表5 方法1適用結果(被験者 B)

アドレス	送信回数	対応候補数	手法1適用結果要素数	全選択時の候補数
1	1059	158	8	3
2	917	157	3	21
3	613	105	8	3
4	555	96	2	34
5	527	98	2	34

同報先の推薦には、日本語入力で用いられる予測変換技術を応用し、推薦候補が多数ある場合には、送信頻度の高いアドレスやアドレスの組合せを利用して、候補の絞り込みを行う方法を用いた。また、アドレスの組合せ候補が多数ある場合に対する絞り込み手法についても提案した。

更に、本手法の有効性を検証するために、実際のメール送信履歴を用いて評価を行い、その結果、ユーザはより少ないアドレス選択回数で正しい送信先を得ることが可能になることを確認し、誤送信の防止効果があるとの見通しを得た。

また、本方式は、メール送信時にもアドレスの組合せの妥当性をチェックし、妥当性が低い場合には、ユーザに正誤の確認を促すことができるので、誤送信の防止効果をさらに高めることができるだろう。

## 参考文献

- [1] 日本商工会議所, 電子メール利用に関するアンケート調査結果について, (<http://www.kentei.ne.jp/email/ankeTo/ankeTo.html>), 2002
- [2] 増井俊之, ペンを用いた高速文章入力手法. In Proceedings of Workshop on Interactive System and software 1997 (WISS 1997), pp.51-60, 日本ソフトウェア科学会, 近代科学社, December 1997
- [3] Tegi Ccommunications, T9 (<http://www.t9.com/japan/>)
- [4] 上田宏高 柳沢豊 塚本昌彦 西尾章治郎, 電子メールの傾向分析への知識獲得手法の適用. 情報処理学会論文誌 Vol.41 No.12 pp3285-3294, Dec. 2000
- [5] JUSTSYSTEM, Shuriken Pro4 (<http://www.justsystem.co.jp/shuriken/product/pro4.html>)
- [6] アスキーソリューションズ, E-メールフィルター, (<http://www.filtering.jp/email/index.html>)