

メール履歴を利用した学習に基づく誤送信メール推定システムの試作

辻野 友孝 伊藤 太樹 柿元 宏晃 加藤 健太

白松 俊 大園 忠親 新谷 虎松

名古屋工業大学大学院工学研究科情報工学専攻

1 はじめに

本稿では、誤送信メールの推定システムを試作する。誤送信メールとは宛先のみを間違えたメールと定義する。本研究では、メールの送信者と受信者の人間関係に着目することにより誤送信メールの推定を行う。人間関係とは上司と部下の上下関係、社内と社外の所属関係、知り合いと友達の友人関係などの関係を指す。部下と上司の様に、人間関係の異なる相手に同じメールを送信しない。同様に、送信者は受信者との人間関係が異なれば、異なる内容のメールを送信すると考えられる。そのため、複数の人間関係を用いる事で、送受信者間の関係とその場合における適切宛先のメールが定義できると考えられる。

人間関係の抽出には送信メール履歴の本文と件名を利用する。送信メールの特徴と送信メール履歴より得られた人間関係の特徴を利用しシステムは送受信者間の人間関係を推定する。システムによって推定された受信側との人間関係と送信メールの宛先との人間関係が異なった場合に誤送信であると判断しユーザに警告文を提示する。

2 既存手法

現在、誤送信を防ぐための様々な方法が存在する。代表的な手法に、添付ファイルを添付した場合に警告文を提示する方法、キーワードマッチングにより重要な内容かどうか判断し警告する方法、異なるドメインにメールを送信する場合警告をする方法などがある。

送信者と一人一人の宛先との人間関係は異なっており、誤送信とは本質的には人間関係の違う相手にメールを送信してしまう事である。しかし、既存の手法ではメールの内容を加味しておらず、人間関係を判断できないため精度良く誤送信メールを推定する事が出来ない。そこで、多くの既存手法では再現率を向上させるため、キーワードマッチングを使いメールの重要度が高い場合にも警告を発するようにしている。そのため実際には誤送信ではない多くのメールに対し警告を表示してしまい、ユーザは警告を軽視してしまうという欠点が生じている。

3 誤送信推定システム

3.1 人間関係を利用

本研究では誤送信推定に人間関係を利用する。送信者と受信者の間には様々な関係がある。例えば上下関係、所属関係、友人関係等がある。相手により人間関係は異なるので複数の人間関係を利用することでユーザを識別できると考えられる。また、所属関係や友人関係は比較的判別しやすい人間関係で、メールアドレスのドメイン名や文章の語尾から判別することができる。しかし上下関係はメールの内容を考慮しなければ非常に推定することが困難であると考えられる。

A System for Estimating Human Error in Sending Email based on Email Histories using Machine Learning
Tomotaka TSUJINO, Taiki Ito, Kakimoto Hitoaki Kenta Kato Shun Shiramatsu, Tadachika OZONO, and Toramatsu SHINTANI
Dept. of Computer Science and Engineering, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya, 466-8555 JAPAN

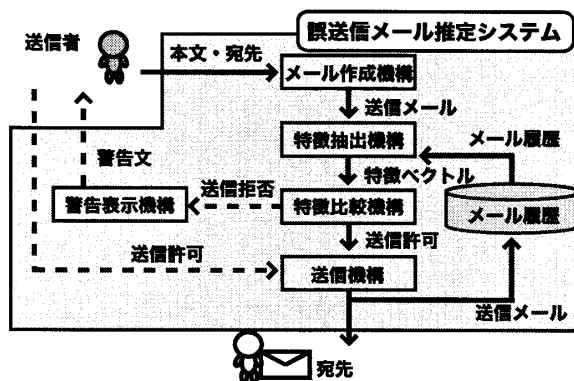


図 1: システム構成

そこで本稿ではシステムが上下関係を示す特徴語の抽出を行う部分について述べる。システムはメールの内容から人間関係の特徴語の抽出を行う。

3.2 システム構成

本研究では学習を利用した誤送信メール推定システムの試作をした。本システムはクライアントサイドで実装を行った。本システムの構成を図 1 に示す。

まず、ユーザはメール作成機構に本文と宛先を入力する。システムは作成した送信メールとメール履歴を特徴抽出機構で MeCab¹ を利用し形態素解析を行い、出現した形態素特徴ベクトルを生成する。作成した特徴ベクトルと、メール履歴から得られた人間関係の特徴ベクトルを特徴比較機構で比較を行い送信者と受信者間の人間関係の推定を行う。システムは推定結果と実際の宛先の人間関係が等しい場合、メールの送信を許可し送信機構によりメールが送信する。人間関係が等しくなく誤送信と推定した場合はシステムは送信を拒否し、警告文表示機構がユーザに警告文を表示する。警告が間違いであった場合、システムはユーザにより送信許可が出されることで送信機構からメールが送信する。送信されたメールはメール履歴に保存され、次回からの推定に利用される。

4 表現の違いに基づく人間関係の推定

4.1 実験

本実験ではメール履歴から送受信者間の人間関係を得ることができるか人間関係を表す特徴語抽出を行った。所属関係のようなメールのドメイン名から簡単に関係が判断できるような関係ではなく、メールの内容を考慮しないと取得できない上下関係の特徴語を抽出した。また、特徴抽出方法の違いによる抽出される特徴語の違いについて検証をした。実験データは大学院生 1 名の研究室 PC で送信したメール 200 通を対象とした。本実験では学生のメールを利用するので「目上クラス」は教員と先輩、「目下クラス」は同級生、後輩とした。メーリングリストのような複数にメールを送信する場合は宛先の中で最も地位の高い人が属するクラスをクラスとした。以上のルールを適用すると、送信履歴 200 通のうち 127 通が目上クラス、73 通が目下クラスとなった。

¹<http://mecab.sourceforge.net>

表 1: 名詞 (tf-idf)

目下クラス (c_1)		目上クラス (c_2)	
単語	tf-idf	単語	tf-idf
例	16.5	辻野	56.7
連絡	12.8	お願い	42.7
メール	12.4	写真	41.1
夏	12.0	3点	32.5
問題	11.8	人	32.4

表 2: bigram (tf-idf)

目下クラス (c_1)		目上クラス (c_2)	
単語	tf-idf	単語	tf-idf
てください	29.9	行目	82.1
して	29.8	します	68.0
行目	27.3	ました	64.0
ました	23.2	して	61.7
よろしく	22.9	辻野です	49.0

特徴語の抽出方法には tf-idf と語の出現確率を用いた 2 通りを用いた。tf-idf 値は次の式で求められる。

$$tf-idf(t, m) = tf(t, m) \times idf(t) \quad (1)$$

これはメール m における語 t についての tf-idf 値を表している。 $tf(t, m)$ はメール m での語 t の出現頻度、 idf は $\log(M/df(t)+1)$ で、 M は全メール数、 $df(t)$ は語 t が含まれるメール数である。

また語の出現確率は次の式で求められる。

$$P(c_i|t) = \frac{df(c_i|t)}{df(t)} \quad (2)$$

これは人間関係のクラス i での語 t についての出現確率を表している。 $df(c_i|t)$ は人間関係クラス c_i における語 t のメールの出現回数である。

また、語句の抽出の際、名詞のみと bigram を利用した場合の 2 パターン行い、抽出する特徴語を比較した。

4.2 結果・考察

事前実験として文献 [1] で述べられているように助詞と助動詞に着目し、tf-idf を利用し特徴語を抽出した。その結果、目上、目下、共にほぼ同じような特徴語が得られた。このことより今回利用するメールはメールの内容を考慮しなければ人間関係を推定することができないデータである事が分かる。本実験ではピリオド、カンマなどの 30 語をストップワードとし特徴語にならないようにした。また、出現確率を利用した方法では同じ人間関係クラスに属するメールの 20 % 以上に出現しない語も特徴語にならないようにした。

実験の結果を表 1, 2, 3 に示す。

表 1 は名詞のみから tf-idf を利用し特徴語を抽出した結果である。目下クラスは「例」「連絡」など説明や相談に用いられる語が抽出された。目上クラスにはフォーマルな形式でメールを書いているため送信者名が抽出された。

表 2 は bigram から tf-idf を利用し特徴語を抽出した結果である目下クラスは「てください」「よろしく」など指示をあたえている表現が抽出された。目上クラスからは「します」「ました」など相手からの指示に受け答えしている表現が抽出された。また、両方のクラスに「行目」「して」という語が特徴語として抽出されている。

表 3: bigram (確率)

目下クラス (c_1)		目上クラス (c_2)	
単語	$P(c_1 t)$	単語	$P(c_2 t)$
よろしく	0.77	3点	1.00
か?	0.58	新谷研	1.00
てください	0.48	計算機	1.00
ている	0.47	よろしくおねがい	0.95
ません	0.41	おねがいし	0.92

表 3 は bigram から出現確率を利用し特徴語を抽出した結果である目下クラスは tf-idf を利用した場合と同様に指示を与える表現、に加え質問表現などが抽出された。目上クラスからは「3点」「新谷研」「計算機」等の毎月の報告の際に使用した表現が抽出された。

これらの結果より、名詞のみを抜き出すよりも bigram を用いて特徴語を抽出した方が指示や応答などの表現を抽出できた。目上、目下関係の抽出には bigram の方が優れていると考えられる。また、tf-idf を用いた場合では良く利用する語は tf 値が高くなるため特徴語として検出されやすく目上、目下両方のクラスで特徴語とされている語がある。それに対し出現確率を用いた方法では片方のクラスにしか出現しない語を抽出している。このことより、tf-idf より出現確率を用いた方がより良い特徴語を抽出できるのではないかと考えられる。

5 おわりに

本稿ではメールの履歴を利用した学習に基づく誤送信メール推定システムについて述べた。本システムは意図した宛先と異なる宛先を指定したメールを検知することを目的としている。送信メールの履歴から各ユーザごとの特徴ベクトルと送信メールの内容と特徴ベクトルより宛先との人間関係を推定する。本研究ではメール内の固有名詞からユーザごとの特徴ベクトルを計算して誤送信メールを推定した。推定された結果の人間関係と実際の宛先との人間関係が異なる場合システムは誤送信と判断し、ユーザに警告を表示する。

本稿ではメール履歴より送受信者間の上下関係の特徴語の抽出を行った。特徴語を抽出する際、名詞と bigram、tf-idf と出現確率をもちいた手法を比較した。今回の実験では bigram から語の出現確率を用いた方法が最も優れた特徴語抽出をすることができた。抽出された特徴語はそれぞれの人間関係に人間関係のある語が抽出された。このことからメールの内容は人間関係の抽出に利用できると考えられる。

今後の課題として本稿では送信者と受信者間の人間関係はあらかじめ既知であるとしたが、送信相手の数が増えると全てを定義するのは非常に困難になる。そこで、文献 [2] の手法を用い、人物に関するキーワードを抽出し自動的に送信相手との人間関係を抽出する方法を取り入れることが考えられる。また、今回の実験では 1 人のメール履歴を用いたが、多くのユーザからの情報を利用して人間関係を推定しようとした場合メールにはプライバシー問題を考慮しなければならない。

参考文献

- [1] 西原陽子, 砂山渡, 谷内田正彦, “発話テキストからの人間の中の良さと上下関係の推定”, 電子情報通信学会論文誌. Vol.J91D, No.1 pp. 77-88
- [2] 松尾豊, 石塚満, “Web からの人物に関するキーワード抽出”, 人工知能学会論文誌. 17 巻 3 号 D (2002 年)