

人間関係を利用した誤送信メール防止システムの試作

A System for Preventing Human Error in Sending Email Using Human Relationship

辻野友孝[†] 白松俊[†] 大園忠親[†] 新谷虎松[†]

Tomotaka Tsujino, Shun Shiramatsu, Tadachika Ozono, Toramatsu Shintani

1 はじめに

近年、メールは簡単に相手に情報を伝達することができるため、日常のコミュニケーション手段として頻繁に利用されている。しかしメールは、宛先を間違える、添付ファイルを間違えるなどの操作のミスで簡単に情報が漏洩してしまう。

本研究ではローカルプロキシサーバ上に誤送信メール防止システムを構築した。本稿における誤送信とは、部下に送るメールを誤って上司に送ってしまったなど、メールの件名や内容が受信者と送信者の関係に適切でないメールを送信することと定義した。本システムはローカルプロキシサーバ上に構築した為、メールサーバに特別なシステムを導入する必要は無い。またユーザは任意のローカルのメーラーを利用できる。本システムは送受信者の両者に導入した場合は暗号化により全ての誤送信を防ぐ。また、送信者のみに導入する場合は人間関係を利用して誤送信を防止する。本稿で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人間関係と呼ぶ。人間関係が異なれば、部下には連絡のメール、上司には報告のメールを送るなど、メールの内容も異なると考えられる。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な宛先のメールが定義できると考えられる。

本システムは送信者側のプロキシサーバで、送信メールの特徴やメールの履歴より得られた人間関係の特徴を利用し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人間関係と送信メールの宛先と送信者の人間関係が異なる場合に誤送信であると判断する。システムは誤送信と判断した場合、ユーザに警告文を提示することでメールの誤送信を防ぐ。

送受信者の両者にシステムを導入した場合は暗号化により誤送信を防止することも可能である。送信者のシステムでメールを暗号化して送信する。受信者側のシステムで復号できなければメールを削除することで受信者にメールを閲覧できないようにした。これにより、受信者側にもシステムが設置してある場合は確実に誤送信を防止できる。

2 誤送信防止手法

2.1 既存システム

送信時にメールの誤送信を防止する手法として、添付ファイルを添付した場合に警告文を提示する手法、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手法、宛先数が一定数以上超えると警告を表示する手法、異なるドメインにメールを送信する場合警告をする手法などがある。また送信後に対応できる手法として、メールを暗号化する手法、メール本文には web サイトのアドレスだけ記載し、本文はそのアドレスにアクセスしないと読めないようにする方法などがある。

多くの既存サービス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。

2.2 人間関係を利用

本研究ではメールの内容に着目した誤送信メールの検出方法を提案する。メールの内容からどの宛先に送るメールなのか推定する場合大きく 2 つの問題が考えられる。1 つ目はメールの本文は短かく、メールの件名や本文中に宛先を特定できる特徴が現れない場合があることである。例えば、「先ほどの件、了解しました」とメールの本文に書いてある場合、誰に送ったのか人間でも判断することができない。2 つ目はメールでは同じ相手と様々な話題でメールをする点である。同じ相手であっても、研究、生活、合宿など様々な話題でメールを送信する。そのため、単純にキーワードマッチングによる手法で相手を特定することは困難である。

本研究では誤送信防止に人間関係を利用する。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。送信者と一人一人の宛先との人間関係は異なっており、人間関係が異なる相手にメールを送信してしまうことが問題である。そこで、宛先への適切さではなく、人間関係クラスへの適切さを利用して誤送信を推定する。メールの本文と内容から宛先を推定するのは困難であるが、宛先がどの人間関係のクラスに属するか推定することは可能であると考えられる。また、人間関係クラスという複数の宛先の集合を利用することで、学習用のデータの収集がそれぞれの宛先ごとにデータを収集するよりも容易になる。

3 誤送信メール防止システム

本システムでは送信メールの件名、本文と送信メール履歴を比較することで誤送信を判定する。本システムでは宛先との人間関係は既知とする。本システムでは送信メールから作成した特徴ベクトルと各人間関係クラスから作成した特徴ベクトルを比較することで相手との人間関係を推定する。システムは推定結果の人間関係と宛先との人間関係が異なる場合誤送信と判定し、ユーザに警告を出すことで誤送信を防止する。今回は研究室で利用しているメールを利用して目上 - 目下関係の誤送信を判定した。研究室内のメールであるため、共通の話題でメールのやり取りを行うのでキーワードマッチングによる方法では誤送信を推定することはできない。また、どちらのクラスにも敬語を利用したメールを送信する。そのため目上 - 目下関係の分類は困難であると考えられる。

3.1 特徴ベクトル

本システムはメールを MeCab¹ を利用し形態素解析を行い人間関係の特徴ベクトルを作成する。形態素解析した結果から bigram を作り *tf-idf* 値を計算し特徴ベクトルを作成した。名詞を【名詞】という文字列に置換することで隠蔽し、話題の違いを吸収した。これにより、文の言葉遣いなどを考慮した。

特徴ベクトルの比較は SVM(Support Vector Machine) を利用した。本システムでは、データマイニングツール Weka² のライブラリを利用する事により行った [1]。Weka は、オープンソースのデータマイニングソフトである。本システムでは SVM に *weka.classifiers.functions.SMO* を用いた。特徴比較機構では、送信履歴の各人間関係の特徴ベクトルを SVM にトレーニングデータとして学習させ、送信メールの特徴ベ

[†]名古屋工業大学大学院 工学研究科 情報工学専攻¹<http://mecab.sourceforge.net>²<http://www.cs.waikato.ac.nz/ml/weka/>

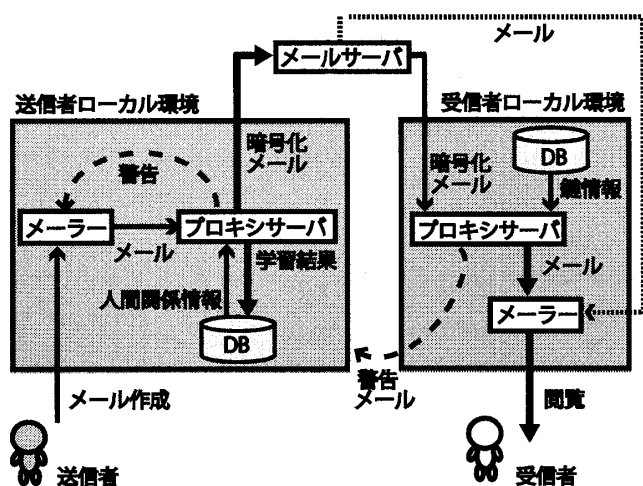


図 1: システム概要図

クトルをテストデータとしてどの人間関係に属するか推定する。

3.2 暗号化

送信者、受信者の両方にシステムを導入することで暗号化により誤送信メールを閲覧できないようにした。既存手法として、暗号化したメールを送信することで誤送信したメールを受信者が復元できず閲覧できないようにする手法がある。本システムでは送信者側のプロキシサーバ上でメールを公開鍵で暗号化し、受信者側のプロキシサーバ上で秘密鍵を使い復号化を行う。復号化が行えなかったメールは誤送信と判断し受信者側のプロキシサーバ上でメールを削除する。その為、受信者に誤送信メールが届いたことを通知しない。

3.3 システム

本研究では人間関係を利用した誤送信メール推定システムの試作をした。本システムの概要を図 1 に示す。本システムはローカルプロキシサーバ上で誤送信推定を行う。プロキシサーバで誤送信でないと判断した場合は太線の矢印、誤送信と判断した場合は破線の矢印の挙動を行う。また、送信者のみにシステムを導入した場合は、メールサーバにメールが届いた後、点線の挙動を行う。

送信者側のみにシステムを構築した場合は人間関係を利用して誤送信推定を行う。まず、送信者はローカル環境にある任意のメーラーでメールを作成し送信する。送信されたメールはローカルに設置してあるプロキシサーバに届く。プロキシサーバ上にあるシステムはメールの件名と本文を抽出する。抽出した文を形態素解析を行い、出現した形態素の *tf-idf* 値から特徴ベクトルを作成する。送信メールから作成した特徴ベクトルと、DB に蓄えられている人間関係の特徴ベクトルを比較する。比較した結果の人間関係クラスが、メールの宛先の人間関係と同一ならばメールを送信し、異なればシステムは誤送信と判断しメーラーに警告を表示する。警告が表示された場合、メールの件名に送信許可を示す文字列を付与することで送信する。また、送信メールの情報を DB に保存し次回からの推定に利用する。

送受信者の両者にシステムを構築した場合は、人間関係による推定を行った後、メールを公開鍵暗号で暗号化しメールサーバに送信する。受信者側のプロキシサーバはメールサーバからメールを取得し受信したメールを秘密鍵を用いて復号する。復号できなかったメールは誤送信と判断しメールを消

去し、警告メールを送信者に送信する。復号できたメールのみを受信者はメーラーで閲覧することができる。

4 複数人の人間関係の利用

本研究では誤送信推定に人間関係を利用する。特定のユーザ間の送信メールの情報を利用するのではなく、上下関係、友人関係など抽象的な関係の情報を誤送信推定に利用する。そのため、他のユーザの学習した人間関係の情報を他のユーザも利用することができるのではないかと考えられる。他のユーザの人間関係の情報を利用することにより、ユーザが今まで送信したことのない人間関係でも誤送信を推定することができるようになる。

しかし複数人の人間関係情報を利用して人間関係の推定を行う際に大きく 2 つの問題が考えられる。1 つ目にプライバシーの問題が挙げられる。メールには他人に見られてはいけない情報が記載されている場合がある。研究室で利用したメールから名詞や固有名詞を記号に置き換え、目上 - 目下という上下関係の 2 クラス分類を行った [2]。結果、名詞を隠蔽し、bigram を利用した場合、83 % の精度で 2 クラス分類できた。また、目上クラスの人に目下クラスに送信するメールを送信してしまう誤送信は 93 % の精度で誤送信を判定することができた。しかし他のアプローチや、他の品詞の置換なども検討する必要がある。

2 つ目に人によって同じ人間関係でも、接し方が違うということが考えられる。目下の人に敬語を使う人もいれば、使わない人もいる。同じ人間関係でも、人によって接し方が違うのでどのように個人差を吸収するのが問題である。そのために重み付けをすることでユーザ自身の送信メール履歴を重視する必要があると考えられる。

5 おわりに

本稿では人間関係に基づく誤送信防止システムについて述べた。システムはローカルプロキシサーバ上に作成した。本システムは 2 種類の誤送信防止機能をもつ。1 つ目が人間関係を利用した誤送信防止方法である。メールの件名と本文を形態素解析し、特徴ベクトルを生成した。送信メールの履歴から各人間関係の特徴ベクトルと送信メールの特徴ベクトルを比較することにより宛先との人間関係を推定する。2 つ目がパスワードを利用した誤送信防止方法である。送信者側でメールを公開鍵を利用して暗号化しを行う。受信者側が受信したメールを秘密鍵で復号化できなかった場合誤送信と判断し、ユーザから閲覧できないようにした。

今後の課題として、送信メール以外の情報の利用が考えられる [3]。送信メールだけでは学習に用いる情報が少ない。その為、ユーザの環境の変化にも対応しにくい。送信メール以外の情報を利用することで、ユーザの環境の変化を随時反映していくことにより推定精度が向上すると考えられる。

参考文献

- [1] Lan Witten, Eibe Frank, "Data Mining: Practical machine learning tools and techniques with Java implementations", Morgan Kaufmann Series in Data Management Systems, June 2005.
- [2] 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, "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", 第 72 回情報処理学会全国大会講演論文集, (2010)
- [3] 上田芳弘, 成田仁志, 加藤直孝, 林克明, 南保英孝, 木村春彦, "テキストマイニングと強化学習を用いた電子メール自動分配", 電子情報通信学会論文誌 D-I Vol.J87-D-I No.10 pp.887-898 2004 年 10 月