多クラス分類による電子メール誤送信検出手法

柴田 秀哉† 加藤 守† 郡 光則† William S. Yerazunis††

† 三菱電機株式会社 情報技術総合研究所 〒 247-8501 神奈川県鎌倉市大船 5-1-1
†† Mitsubishi Electric Research Laboratories 201 Broadway, Cambridge, MA 02139, USA
E-mail: †{Shibata.Hideya@cb, Kato.Mamoru@dn, Kori.Mitsunori@ab}.MitsubishiElectric.co.jp,
††yerazunis@merl.com

あらまし 本稿では，多クラス分類による電子メールの誤送信検出手法を提案する．電子メールの誤送信による情報漏洩事故が後を絶たない中，人手による情報管理負担の低減を目的とした誤送信自動検出技術が求められている．電子メールには，想定される宛先が付随するため，宛先を分類クラスに持つような電子メールの多クラス分類問題を解くことで，分類結果と実際の宛先との差分を誤送信として検出することが可能となる．本稿では，誤送信検出のための多クラス分類問題を定式化し，機械学習を用いた誤送信検出手法を提案する．また，業務メールデータ約 600 万件を用いた評価実験により，提案手法の有効性を実証する．

キーワード 電子メール誤送信検出，テキスト分類，機械学習

A Method of Missent E-mail Detection Based on Multi-Class Classification

Hideya SHIBATA†, Mamoru KATO†, Mitsunori KORI†, and William S. YERAZUNIS††

† Information Technology R&D Center, Mitsubishi Electric Corporation
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501 Japan
†† Mitsubishi Electric Research Laboratories 201 Broadway, Cambridge, MA 02139, USA
E-mail: †{Shibata.Hideya@cb, Kato.Mamoru@dn, Kori.Mitsunori@ab}.MitsubishiElectric.co.jp,
††yerazunis@merl.com

Abstract The recent information explosion has often caused information leakage via e-mails, and has increased the necessity of the technology which enables the missent e-mails detection. One of the ways to detect missendings is to solve a multi-class e-mail classification problem which defines the intended addresses of e-mails as the classes, and to find the differences between the results and the actual addresses. In this paper, we formulate a multi-class classification problem to detect missent e-mails, and propose a detection method using machine learning. We also show the effectiveness of our method through the evaluation with about six million business e-mails.

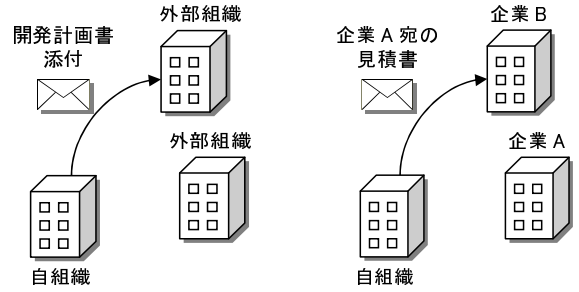
Key words missent e-mail detection, text classification, machine learning

1. はじめに

近年，情報量の増加が著しく，人手による情報管理のみでは誤りによる情報漏洩を防止できないという問題がある．情報漏洩の 1 つの形態として，電子メールの誤送信がある．日本ネットワークセキュリティ協会が出した調査報告によると，2009 年に起きた情報漏洩事故は 1,539 件に上り，電子メールによる事故はその 7%を占めている [1]．このような背景があり，人手による情報管理負担の低減を目的として，電子メールの誤送信を高精度に自動検出するための技術が求められている．

電子メール誤送信を自動検出するための 1 つの方法として，クラス分類を利用する方法がある．これは，スパムメール検出の分野において良く知られた方法である．ベイズ分類器やサポートベクトルマシン（SVM）などを利用した学習型のテキスト分類器により，高精度なスパムメール検出が可能であることが報告されている [2], [3]．また，筆者らは学習型のテキスト分類器を機密情報の自動判別へ適用する研究開発に取り組んでいる．その成果として，企業における外部送信が許されない機密メールと外部送信可能な非機密メールとを，電子メールの本文情報に基づいて高精度に分類可能であることを評価実験により示した [4], [5]．これにより，機密と判定された外部送信メールを，機密メールの誤送信として検出することが可能となる．

機密メールの外部送信は電子メール誤送信の一種であるが，その他にも検出すべき誤送信が存在する．それは，外部の特定宛先に送信したい電子メールの宛先誤りである．例えば，図 1-(b) に示すように，ある企業 A 宛の見積書を誤って企業 B へ送信すれば，これは一種の情報漏洩である．従って，このような誤送信を検出するための仕組みが必要となる．



(a) 機密メールの外部誤送信　　(b) 外部送信メールの宛先誤り

図 1　電子メール誤送信の形態

本稿では，図 1 で示すように，電子メール誤送信を

- 機密メールの外部誤送信
- 外部送信メールの宛先誤り

の 2 種類に分けて扱い，これら誤送信を検出可能とするようなクラス分類問題を定式化する．また，機械学習を用いた電子メール誤送信検出手法を提案し，外部送信メールの宛先誤りが高精度に検出可能であることを，業務メールデータ約 600 万件を用いた評価実験により示す．

本稿の構成は以下のとおりである．まず，2 節で先行技術との関係性について述べ，3 節で，電子メール誤送信検出問題を多クラス分類問題として定式化する．4 節では，3 節で定式化した問題に対して機械学習アルゴリズムを適用し，電子メール誤送信を実際に検出する方法を述べる．5 節では，提案手法に対する評価結果を報告する．最後に，6 節で本稿のまとめを行う．

2. 先行技術との関わり

電子メール誤送信の検出手法としては，ポリシー設定によるルールベースの検出が良く知られる。例えば，頻繁に電子メールをやり取りする取引先等のドメイン情報と組織名を予めシステムに登録しておき，電子メール送付先のドメイン情報と本文に記述された組織名とが一致しない場合に，これを誤送信として検出するというものである．

ポリシー設定による手法は，ユーザへのルール強制によるユーザ負担増加，柔軟なポリシー設定の困難性など，多くの課題を持つ．そこで，筆者らは機械学習に基づく電子メール誤送信検出の研究開発を進めてきた．具体的には，学習型のテキスト分類器により組織内部向けの電子メールを機密メールとして高精度に自動判別可能なことを示した [4], [5]．同時に，システムの運用性向上を目的とし，機械学習を利用する際に課題となる訓練用データ収集の自動化手法を開発した [5].

機密メールの自動判別は，スパムメール検出と同様に 2 クラスのテキスト分類問題として捉えることができ，既存の様々な分類アルゴリズムを適用してこれらの問題を解くことができる．機密メール誤送信検出がスパムメール検出と性質を異にしている点は，分類器の判定結果が検出すべき電子メールを直接的には決定しない点である．スパムメール検出は,「検出すべき電子メール」と「検出すべきでない電子メール」との 2 クラス分類であり，この意味で検出処理と分類処理は同一である．一方，機密メール誤送信検出においては，電子メールの内容が同一であっても，その宛先によって電子メールを検出すべきか否かが異なる．従って，機密メール誤送信検出では「検出すべき／検出すべきでない」という観点ではなく,「外部送信可能／外部送信不可能」という観点で電子メールを分類する必要がある．その上で，分類器の判定結果と宛先との差異を誤送信として検出する，という検出処理が必要となる．

外部送信メールの宛先誤り検出についても，機密メール誤送信検出の考え方を拡張し適用することができる．すなわち，電子メールの想定する宛先を 1 つのクラスとするような多クラス分類を考えることで，分類器による判定結果と実際の宛先との差異を誤送信として検出することができる．こうして，機密メール誤送信検出と外部送信メールの宛先誤り検出を，クラス分類問題として包括的に扱うことが可能となる．

注意すべき点として，電子メール誤送信検出においては，同報メールの存在を考慮に入れなければならない．すなわち，電子メールが複数の宛先に同時に送信されることを想定している場合，当該電子メールが属する正解クラスを 1 つに定めることはできない．従って，電子メール誤送信検出問題を多クラス分類問題と捉える際には，複数クラスへ属することを許容するような分類モデルが必要である．

3. 電子メール誤送信検出問題の定式化

本節では，電子メール誤送信検出問題をクラス分類問題として定式化する．なお，以下では電子メール誤送信検出の実施を想定している環境を自組織と表現し，その他の組織を外部組織と呼ぶ．

3.1 組織の定義

まず，電子メールアドレスの観点から組織を定義し，考えている問題における組織を定式化する．全ての電子メールアドレスからなる集合を $\mathcal{A}$（address の頭文字）で表し，集合 $\mathcal{A}$ の任意の部分集合を組織と定義する．

この定義の下で，自組織に対応する組織を A_0 とおく．また，n 個の互いに交わらない外部組織 A_i $(i = 1, \cdots, n)$，および A_0〜A_n に含まれない電子メールアドレスの集合 A_ω を考える．すなわち，

$$A_\omega := \mathcal{A} \setminus \left(\bigcup_{i=0}^{n} A_i \right) \tag{1}$$

である．ここでは，“ω”を「任意の自然数と異なる記号」という意味で便宜的に用いている．以下では，A_ω も 1 つの外部組織として扱う．更に，添え字集合 $\mathcal{I}$（index の頭文字）を

$$\mathcal{I} := \{0,\ 1,\ \cdots,\ n,\ \omega\} \tag{2}$$

で定義しておく．

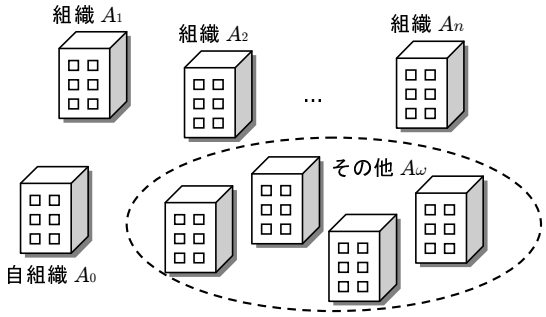


図 2　組織の関係

この状況を図 2 に示す．自組織 A_0 としてある企業を考えている場合，図 2 の状況は，n 個の主要取引先 A_i $(i = 1, \cdots, n)$ とその他の外部組織からなる集合 A_ω と捉えることができる．最終的に実現したいことは，例えば外部組織 A_1 宛の電子メールを誤って外部組織 A_2 へ誤送信したとき，これを正しく検出することである．

3.2　クラスの定義

次に，電子メール誤送信検出問題に使用する分類クラスを定義する．分類したいのは電子メールであるので，当然，クラスは電子メールからなる集合として定義される．そこで，自組織 A_0 から送信される全ての電子メールからなる集合を $\mathcal{M}$（mail の頭文字）で表し，集合 $\mathcal{M}$ の部分集合としてクラス C_i $(i \in \mathcal{I})$ をそれぞれ

$$C_0 := \{\, m \in \mathcal{M} \mid m \text{ は外部組織へ送信不可} \,\}, \tag{3}$$

$$C_i := \{\, m \in \mathcal{M} \mid m \text{ は外部組織 } A_i \text{ へ送信可} \,\} \quad (i = 1, \cdots, n, \omega) \tag{4}$$

で定義する．クラス C_i は組織 A_i に対応している．クラス C_0 は，自組織という特別な組織に対応するクラスであるため，他のクラスと定義が異なることに注意されたい．

定義より，明らかに

$$\mathcal{M} = \bigcup_{i \in \mathcal{I}} C_i \tag{5}$$

である．また，任意の $i \neq 0$ に対して

$$C_0 \cap C_i = \emptyset \tag{6}$$

が成立する．しかしながら，一般には，相異なる $i,\ j \in \mathcal{I}$ に対して，$C_i \cap C_j = \emptyset$ が成立するとは限らない．電子メール m が $m \in C_i \cap C_j$ を満たすことは，外部組織 $A_i,\ A_j$ の両方に m を送信可能であることを意味しており，この定義が同報メールの存在を許容していることを表している．

3.3　クラス分類問題への定式化

2 節で述べたように，同報メールに対しては正解クラスが 1 つに定まらないため，$C_0, \cdots C_n, C_\omega$ の $(n+2)$ クラスからなる多クラス分類を単純に考えることはできない．ここでは，$(n+2)$ 個の 2 クラス分類問題を解くことで，複数クラスへ属することを許容した多クラス分類を実現する．

各クラス C_i $(i \in \mathcal{I})$ に対して，集合 $N_i \subset \mathcal{M}$ を

$$N_i := \mathcal{M} \setminus C_i \tag{7}$$

で定義する．すなわち，

$$N_0 = \{\, m \in \mathcal{M} \mid m \text{ は外部組織へ送信可} \,\}, \tag{8}$$

$$N_i = \{\, m \in \mathcal{M} \mid m \text{ は外部組織 } A_i \text{ へ送信不可} \,\} \quad (i = 1, \cdots, n, \omega) \tag{9}$$

である．定義より，任意の $i \in \mathcal{I}$ に対して

$$C_i \cap N_i = \emptyset \tag{10}$$

が成立する．文献 [4], [5] における機密メールの自動判別は，C_0 と N_0 による 2 クラス分類に相当する．本稿で扱う電子メール誤送信検出では，各 $i \in \mathcal{I}$ に対して，C_i と N_i による 2 クラス分類を実行する．この状況を図 3 に示す．図中において，点線で囲まれた網掛け部分が N_i を表している．

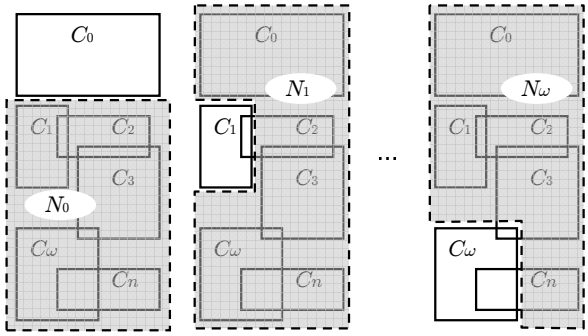


図 3　誤送信検出のためのクラス分類問題

なお，本稿では各 2 クラス分類問題は互いに独立であると仮定する．従って，実質的には複数の 2 クラス分類問題を個別に扱っているに過ぎず，改良の余地が残されている．例えば，クラス間の相関を考慮に入れ，判定結果を補正するなどの改善策が考えられ，今後の課題である．

以上の設定の下で，電子メール誤送信検出問題における識別関数 f を，

$$f : \mathcal{M} \to \mathcal{P}(\mathcal{I}) \tag{11}$$

の形で与えることができる．但し，$\mathcal{P}(\cdot)$ は集合の冪集合を表す記号である．すなわち，識別関数 f が理想的な場合，電子メール $m \in \mathcal{M}$ と添え字 $i \in \mathcal{I}$ に対して，関数 f は

$$i \in f(m) \quad \Longrightarrow \quad m \in C_i \tag{12}$$

という判定を与える．但し，式 (3), (4) から，$0 \in f(m)$ であった場合，$f(m) = \{0\}$ でなければならない．

通常，識別関数の終集合は添え字集合 $\mathcal{I}$ であるが，ここでは複数クラスへ属することを許容するため，終集合が冪集合 $\mathcal{P}(\mathcal{I})$ となる．例えば，ある電子メール $m \in \mathcal{M}$ に対して $f(m) = \{3,\ 5\}$ であった場合，これは m を送信しても良い外部組織が A_3，A_5 の 2 つであると判定されたことを意味する．

3.4 判定誤りの種類

電子メール誤送信検出問題における判定誤り，およびその危険性について記述する．判定誤りには，検出漏れと過剰検出の 2 種類が存在する．以下，それぞれについて説明する．

- 検出漏れ

クラス C_i に属さない電子メール m に対し，誤って $i \in f(m)$ と判定する場合が該当する．この状況を図 4 に示す．このとき，仮に電子メール m を組織 A_i へ誤送信した場合，この誤送信を検出する方法は存在しない．従って，この種の誤りは情報漏洩事故へ直結する．電子メール m を組織 A_i 以外のある組織 A_j へ送信する場合，ここでの判定誤りは影響しない．

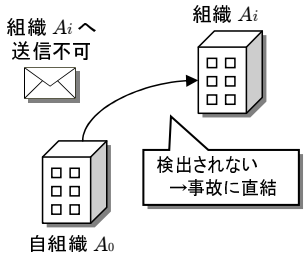


図 4 検 出 漏 れ

- 過剰検出

クラス C_i に属する電子メール m に対し，誤って $i \notin f(m)$ と判定する場合が該当する．この状況を図 5 に示す．このとき，電子メール m を組織 A_i に正しく送信した場合，誤判定により検出されるが，人手により正しい送信であることが確認できる．従って，この種の誤りは人手の負担を増大させるが，情報漏洩事故へは直結しない．電子メール m を組織 A_i 以外のある組織 A_j へ送信する場合，ここでの判定誤りは影響しない．

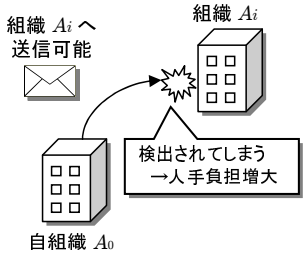


図 5 過 剰 検 出

以上の議論から，検出漏れと過剰検出の低減はトレードオフの関係にあり，電子メール誤送信検出においては，検出漏れの低減が優先される．

4. 電子メール誤送信検出手法

本節では，前節で定式化した電子メール誤送信検出問題を，具体的に解くための手法を提案する．

4.1 前 処 理

3.3 節で定式化した問題では，各 $i \in \mathcal{I}$ に対して，C_i と N_i による $(n+2)$ 個の 2 クラス分類を実行する．多くの場合，集合 C_i は単一組織宛の電子メール集合であり，集合 N_i はその補集合である．従って，集合 N_i のサイズは集合 C_i と比較して大きく，かつ多様な話題の電子メールを含む．このことは，分類精度を下げる一因となり得る．そこで，集合 N_i に代わる，よりサイズが小さい集合を導入する．

式 (6), (7) より，集合 N_i に対して

$$C_0 \subset N_i \quad (i = 1,\ \cdots,\ n,\ \omega) \tag{13}$$

が成立する．また，C_0 と N_0 の 2 クラス分類を実施することにより，C_0 に属する電子メールは判定可能である．従って，$i = 1,\ \cdots,\ n,\ \omega$ において，N_i から C_0 を取り除いても，所望の分類が可能である．そこで，集合 $\tilde{N}_i$ を

$$\tilde{N}_i := N_i \setminus C_0\ (= N_i \cap N_0) \quad (i \in \mathcal{I}) \tag{14}$$

で定義し，各 $i \in \mathcal{I}$ に対して，C_i と $\tilde{N}_i$ による $(n+2)$ 個の 2 クラス分類を実行するようにする．当然，$N_0 = \tilde{N}_0$ である．こうすることで，集合 N_i のサイズを小さくするができる．

集合 C_0 は外部送信が許されない機密メールの集合であった．従って，各 N_i $(i \neq 0)$ から C_0 を取り除くことは，実質的に，機密メール外部誤送信検出と外部送信メールの宛先誤り検出とを分離し，2 段階で電子メール誤送信検出を実現することを意味する．具体的な検出手順は次項で述べる．

4.2 検 出 手 順

電子メール誤送信検出の具体的な手順を述べる．ここでは，各 $i \in \mathcal{I}$ に対して C_i と $\tilde{N}_i$ による $(n+2)$ 個の 2 クラス分類を実現する識別関数 $f_i : \mathcal{M} \to \{0,\ 1\}$ が与えられていると仮定する．但し，識別関数 f_i が理想的な場合，電子メール $m \in \mathcal{M}$ に対して関数 f_i は

$$\begin{cases} f_i(m) = 0 \quad \Longrightarrow \quad m \in \tilde{N}_i \\ f_i(m) = 1 \quad \Longrightarrow \quad m \in C_i \end{cases} \tag{15}$$

という判定を与えるものとする．

このとき，電子メール $m \in \mathcal{M}$ に対する誤送信検出手順を以下で与える．以下では，代入操作を“ $\leftarrow$ ”で表す．

step1. $f_0(m) = 0$ のとき，$f(m) \leftarrow \emptyset$ として step2 へ進む．$f_0(m) = 1$ のとき，$f(m) \leftarrow \{0\}$ として step3 へ進む．

step2. 各 $i \in \{1,\ \cdots,\ n,\ \omega\}$ に対して，$f_i(m) = 1$ のとき，$f(m) \leftarrow f(m) \cup \{i\}$ とする．全ての i に対して実行した後，step3 へ進む．

step3. $f(m)$ と実際の m の宛先とを比較し，$f(m)$ に含まれない宛先分を誤送信として検出する．

以上の手順で，電子メール誤送信検出が実現される．step1 により，電子メールが外部送信可能か否かを判定する．これは，機密メールの外部誤送信検出に相当し，文献 [4], [5] において記述されている検出処理に該当する．次いで，step2 により，外部送信メールを想定される宛先毎に分類する．これにより，外部送信メールの宛先誤り検出が可能となる．

上記の検出手順においては，処理の対象を，実際に外部組織へ送信された電子メールのみに制限することができる．何故な

ら，外部組織へ送信されない電子メールは，その内容に関わらず，情報漏洩の危険性を持たないためである．従って，このような電子メールを検出する必要はなく，処理が不要となる．

4.3 識別関数 f_i の生成

各 $i \in \mathcal{I}$ に対して，識別関数 $f_i : \mathcal{M} \to \{0,\ 1\}$ を適当な方法で用意すれば，4.2 節で示した手順により，電子メールの誤送信検出が実現される．ここでは，機械学習を利用した識別関数 f_i の生成について述べる．

機械学習を用いたテキスト分類器を利用する際，アルゴリズムの形態に関わらず，実用上，最も課題となるのが訓練用データの収集手順である．特に，多くの企業や団体にとって電子メールデータは機密情報であるため，一部の管理者以外はデータを閲覧することすら許可されない．そのため，一般の構成員や技術者によるシステム構築・運用が極めて困難となる．

この課題を解消するため，筆者らは，機密メール外部誤送信検出のための訓練用データ自動収集手法を開発した [5]．この手法は，過去に送信済みの電子メールに対して，電子メールアドレスによりラベル付けを行うことにより，自動で訓練用データを収集するというものである．以下に自動収集手順の概要を記す．詳しくは文献 [5] を参照されたい．

step1． 過去に送信済みの電子メールヘッダに記述された電子メールアドレスを元に，訓練用データのラベル付けを実施する．
step2． 誤送信メールを訓練用データとして採用することを避けるため，過去の判定結果と電子メールアドレスによるラベル付けとが一致しないものを訓練用データから除外する．但し，ここでは過去に送信済みの電子メールにはアドレスとは別の何らかの手段でクラス分類が実施されているものとする（[5]，候補抽出方式 3.4）．
step3． 学習の偏りを防ぐため，訓練用データ件数におけるクラス間の偏りを抑制する処理を施す（[5]，選定方式 3.5）．

以上の手順により，識別関数 f_0 を自動生成することができる．この考え方は，$i = 1,\ \cdots,\ n,\ \omega$ に関してもそのまま適用することが可能である．すなわち，step1 において，To や Cc といったヘッダフィールドに外部組織 A_i の電子メールアドレスが含まれるものをクラス C_i へラベル付けし，そうでない電子メールを $\tilde{N}_i$ へラベル付けする．

ここで問題となるのは，step2 における事前の判定結果を如何にして与えるかということである．これはシステムの初期運用時に問題となる．文献 [5] では，この問題をキーワード検索による機密検出の併用によって回避している．例えば，「社外秘」や「システム開発計画書」などの単語を機密情報を表すキーワードとして設定することで，クラス C_0 に属する電子メールの判定を行い，訓練用データのラベル付けにおいてもこの判定結果を利用している．しかしながら，外部組織 $(i = 1,\ \cdots,\ n,\ \omega)$ に関する 2 クラス分類において，キーワード検索を併用することは困難である．何故なら，自組織の機密情報に特徴的なキーワードを自ら設定することと比較し，外部組織に特徴的なキーワードを設定することは容易ではないためである．

現状では，step2 の手順を無視して，初期運用時には step1 と step3 による訓練用データ収集を実施するという単純な回避策を取らざるを得ない状況である．この点の改良は今後の課題である．

5. 評　　価

本節では，前節までで述べた提案手法を用いて，外部送信メールの宛先誤り検出に関する評価を実施したので，その結果を報告する．

5.1 評価の方針

評価実験は，電子メールアーカイブに対して，提案手法である電子メール誤送信検出手法を適用するという形態で実施する．電子メールアーカイブを実運用に近い形で動作させ，電子メールの誤送信検出精度を調べる．本評価では，外部送信メールの宛先誤り検出に関する精度のみを評価対象とし，機密メールの外部誤送信検出に関する評価は実施しない．

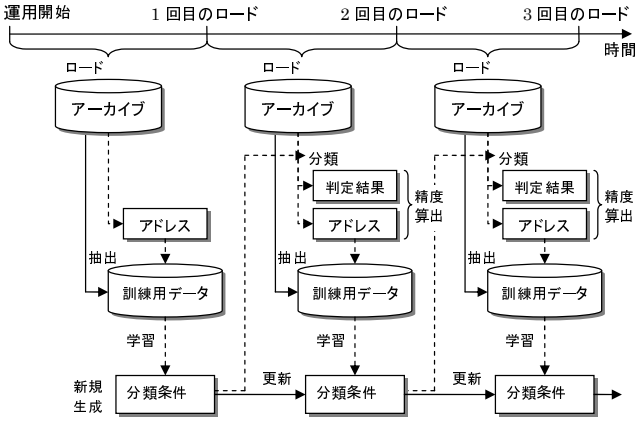


図 6　電子メールアーカイブ運用イメージ

評価実験で使用する電子メールアーカイブの運用イメージを図 6 に示す．電子メールアーカイブのロード周期に併せて，1 周期分の電子メールに対して誤送信検出のための分類を実施し，分類精度を算出する．分類後，分類対象であった電子メールを対象に訓練用データ収集・学習を実施し，分類条件を随時更新する．

評価で用いる訓練用データ収集方法は，4.3 節で示した手順のうち，step1 と step3 を実施することにより実現される方法であるとする．評価データは，実業務で使用した電子メールであり，いずれも適切な宛先へ送信された実績があるため，電子メールの宛先を正解クラスとして訓練用データのラベル付けを実施しても，評価上は問題ない．

なお，本評価により得られる数値は，仮に評価データを誤送信した場合に，どの程度の割合でそれらを検出できるかを表すものであり，過去にどの程度誤送信が発生していたかを調べるものではない．

5.2 評価内容・条件

評価実験の実施条件を以下に記す．

- **評価データ**

使用する評価データは 2009 年 11 月から 2010 年 9 月の約 2

表 1　評価データ内訳

添え字 i	C_i の件数	$\tilde{N}_i$ の件数
1	80,070 件	5,726,499 件
2	52,374 件	5,754,195 件
3	67,099 件	5,739,470 件
4	39,046 件	5,767,523 件
5	25,245 件	5,781,324 件
6	22,067 件	5,784,502 件
7	33,685 件	5,772,884 件
8	16,800 件	5,789,769 件
9	20,056 件	5,786,513 件
10	8,481 件	5,798,088 件
11	5,324 件	5,801,245 件
12	12,772 件	5,793,797 件
13	7,216 件	5,799,353 件
14	5,491 件	5,801,078 件
15	8,550 件	5,798,019 件
16	7,754 件	5,798,815 件
17	9,020 件	5,797,549 件
ω	5,597,318 件	209,251 件

年間に外部送信された業務メール約 600 万件である．評価データの正解クラスを実際の宛先により定義し，本文情報を用いた分類器により，正解の宛先をどの程度再現できるか評価する．

評価で使用する外部組織として，主要な取引先を中心に 17 個の組織を設定した．クラス毎のデータ件数を表 1 に示す．

- **テキスト分類器**

学習型のテキスト分類器として，文献 [4] に記載の種々の分類器の中から，高速であり，分類条件ファイルサイズが学習件数に依らず固定であることを理由とし，Orthogonal Sparse Bigram 方式を適用したベイズ分類器，および Bit Entropy 分類器の 2 種類の逐次学習型分類器を採用し，併用する．併用に際しては，検出漏れの低減を優先するため，どちらか一方の分類器が組織 A_i へ送信不可と判定した電子メールを，総合的に組織 A_i へ送信不可と判定する．

- **ロード周期**

本評価では，電子メールアーカイブのロード周期を電子メール 1 万件分とし，約 600 回の電子メール分類処理を実施するものとする．

- **評価指標**

分類精度の評価指標として，仮に誤送信を起こしたときに検出したい電子メールをどの程度再現できるか，という観点から，検出漏れに関する指標として再現率（recall）を，過剰検出に関する指標として過剰検出率（fallout）を採用する．これらは，テキスト分類，情報検索において一般的に用いられる指標である [6]．但し，過剰検出率に関しては fallout に対する定訳が存在しないため，このように表現する．具体的には，各 $i \in \{1, \cdots, n, \omega\}$ に対して，R_i, F_i を

$$R_i = \frac{\text{正しく } A_i \text{ へ送信不可と判定した件数}}{A_i \text{ へ送信不可な件数}} \tag{16}$$

$$F_i = \frac{\text{誤って } A_i \text{ へ送信不可と判定した件数}}{A_i \text{ へ送信可能な件数}} \tag{17}$$

で定義する．R_i は再現率に，F_i は過剰検出率にそれぞれ対応している．また，全クラスを統合した評価指標として，R, F を

$$R = \frac{\sum_i(\text{正しく } A_i \text{ へ送信不可と判定した件数})}{\sum_i(A_i \text{ へ送信不可な件数})} \tag{18}$$

$$F = \frac{\sum_i(\text{誤って } A_i \text{ へ送信不可と判定した件数})}{\sum_i(A_i \text{ へ送信可能な件数})} \tag{19}$$

で定義する．

表 2　全評価データに対する分類精度

添え字 i	再現率 R_i	過剰検出率 F_i
1	99.3%	11.4%
2	99.6%	5.1%
3	99.6%	4.5%
4	99.4%	7.6%
5	99.7%	5.7%
6	99.7%	8.1%
7	99.5%	9.7%
8	99.7%	4.8%
9	99.7%	11.2%
10	99.7%	11.4%
11	99.7%	3.9%
12	99.6%	19.5%
13	99.7%	13.9%
14	99.7%	10.3%
15	99.6%	23.8%
16	99.7%	19.1%
17	99.7%	12.2%
ω	98.2%	5.1%
全体	$R = 99.6\%$	$F = 5.4\%$

5.3　評 価 結 果

全評価データに対する分類処理が完了した時点の分類精度を表 2 に示す．但し，表中の数値は，全評価データ約 600 万件に対して式 (16)〜(19) を適用して算出した値である．

18 クラス全体での分類精度は再現率 99.6%，過剰検出率 5.4%であり，全体としては，電子メールの誤送信を高い精度で検出可能であることが分かる．しかしながら，クラス別で見ると分類精度にはばらつきがある．特に，C_i に属する電子メール件数が少ないようなクラスにおいて，過剰検出が増える傾向にあることが表 1, 2 より分かる．電子メール件数と過剰検出率との相関を図 7 に示す．

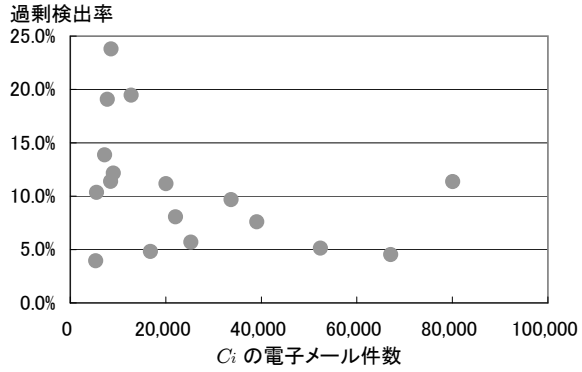
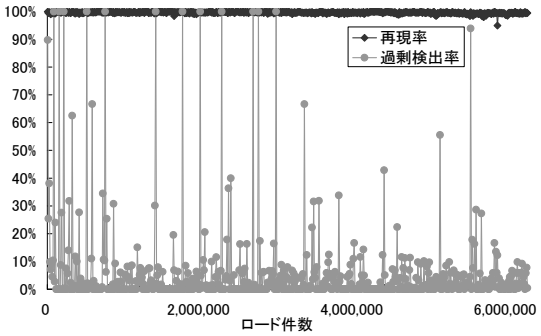


図 7　電子メール件数と過剰検出率との相関

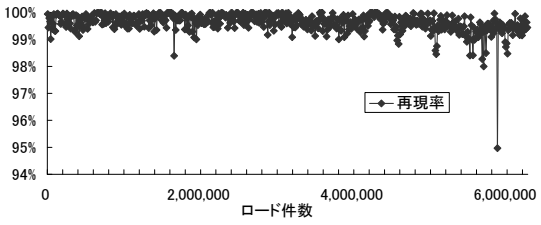
この現象は，訓練用データ件数の偏りが分類精度に悪影響を与えることを表している．4.1 節でも述べたように，提案方式においては，C_i に属する電子メール件数が $\tilde{N}_i$ と比較して小さ

くなる傾向があり，そのため，訓練用データ件数の偏りは本質的な課題と言える．偏りを抑えるための方式（4.3 節，step3）を採用してはいるものの，偏りが一定値を超えるとその影響は無視できなくなる．従って，訓練用データ件数の偏りに影響を受けにくくなるよう，分類器を改良するなどの対策が必要であり，今後の課題である．

続いて，電子メールアーカイブのロード周期に併せて算出した分類精度の時間変化を図 8, 9 に示す．ここでは，代表として組織 A_3 に関するロード周期毎の分類精度を示している．図 8 は，各時点においてロードされた 1 周期分の電子メールに対する分類精度の変遷を表している．一方，図 9 は，各時点で既に分類処理が完了した全電子メールに対する累積の分類精度を表している．各図の (a) には再現率と過剰検出率の両方をプロットし，(b) には変化の小さい再現率のみを拡大して図示している．なお，図の横軸は処理した電子メール件数であり，1 万件毎に点をプロットしている．



(a) 再現率および過剰検出率



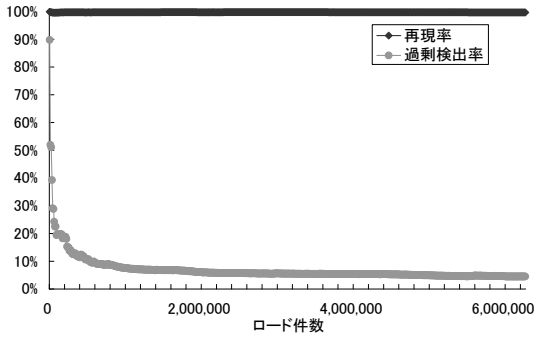
(b) 再現率拡大

図 8　組織 A_3 に関する分類精度率の時間変化（各周）

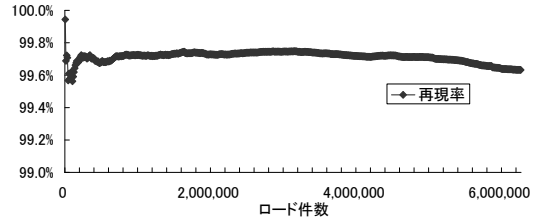
図 8 より，特に過剰検出率において，精度の低い点が散見される．このような点では，過去に例のない内容の電子メールを送信することで誤った判定が為されていると推測される．但し，このような点が連続して発生していないことから，1 周期後にはこれらの電子メールの学習結果が反映され，正しい判定が為されるようになっていることも分かる．このように，局所的に見ると精度が低下する点が散在するものの，全体としては，ある程度高い精度に収束していく様子が図 9 より確認できる．

6. おわりに

本稿では，電子メールの誤送信検出を多クラス分類問題として定式化した．また，機械学習を用いたテキスト分類器を適用し，具体的に誤送信を検出する手法を提案した．提案手法の有効性を示すため，業務メールデータ約 600 万件を使用した評価



(a) 再現率および過剰検出率



(b) 再現率拡大

図 9　組織 A_3 に関する分類精度の時間変化（累積）

実験を実施し，誤送信の一種である外部送信メールの宛先誤りを高精度に検出可能であることを実証した．機密メールの外部誤送信を高い精度で検出できることは，既に文献 [4], [5] で示しており，本稿で示した評価結果と併せて，電子メール誤送信を高精度に検出するための仕組みが整ったと言える．

しかしながら，提案手法にはまだ課題が残されている．例えば，3.3 節で述べたように，各 2 クラス分類を独立な問題としてではなく，何らかの相関を持った問題として定式化することで，より精度の高い分類を実現できる可能性がある．また，4.3 節で述べたように，システムの初期運用時に正確な訓練用データ自動収集に関しても，まだ改良の余地が残されている．

これらの点を踏まえ，今後は，電子メール誤送信検出機能を備えたシステムの運用性向上，更なる高精度化などの課題に取り組んでいく予定である．

文　献

[1] 日本ネットワークセキュリティ協会, “ 2009 年情報セキュリティインシデントに関する調査報告書 ”, http://www.jnsa.org/, Sept. 2010.

[2] P. Graham, “Better Bayesian Filtering”, http://www.paulgraham.com/better.html, Jan. 2003.

[3] M. Kato, J. Langeway, Y. Wu and W.S. Yerazunis, “Three NonBayesian Methods of Spam Filtration: CRM114 at TREC 2007”, Proc. The 6th Text REtrieval Conference, Gaithersburg, Maryland, USA, Nov. 2007.

[4] W.S. Yerazunis, M. Kato, M. Kori, H. Shibata and K. Hackenberg, “Keeping the Good Stuff In: Confidential Information Firewalling with the CRM114 Spam Filter & Text Classifier”, White Paper for Black Hat USA 2010, Las Vegas, Nevada, USA, 2010.

[5] 柴田秀哉，加藤守，郡光則，W.S. Yerazunis, “ 機密メール検出における訓練用データ自動収集手法 ”, DEIM Forum 2010 B4-1, Mar. 2010.

[6] D. Lewis, “Evaluating Text Categorization”, In Proceedings of the Speech and Natural Language Workshop, pp.312-318, 1991.