ベイジアンフィルタリングを用いた迷惑メール対策における多言語環境でのコーパス分離手法の提案と評価

岩永 学[†,☆] 田端 利宏[††,☆☆] 櫻井 幸一[††]

迷惑メールに対する，ベイズ理論を用いた統計的なフィルタリング（いわゆるベイジアンフィルタリング）の研究は以前から行われていたが，2002年に発表されたGrahamの「A plan for spam」[1]をきっかけとしてベイジアンフィルタリングを用いた迷惑メールフィルタが多数開発されるようになった．統計的フィルタリングの場合，複数の言語の電子メールが混在する電子メール環境においては，従来，電子メールの言語ごとにコーパス（学習データ）を分けることが行われてきた．本論文では，日本語と英語のように複数の言語の電子メールが混在する電子メール環境における迷惑メールと正当な電子メールの分類精度向上を目的として，トークンごとに使用するコーパスを選択する方式を提案する．また，従来のメールごとにコーパスを選択する手法を実験により比較し，提案方式の有用性を示す．

Proposal and Evaluation for Improvement of Corpus Separation in Bayesian Spam Filtering on Multi-lingual Environment

Manabu Iwanaga,[†,☆] Toshihiro Tabata[††,☆☆] and Kouichi Sakurai[††]

Statistical filtering using Bayes theory, called Bayesian filtering, is studied for years, and after Graham published an essay "A plan for spam" [1], many implementations of Bayesian filtering have developed. In multi-lingual email environment, which more than one language is used in incoming email, corpus for statistical filtering is usually separated into ones specified to each language. In this paper, we propose a new method in which a corpus is chosen for each token, and then we show the efficiency of our proposed method by experiments in comparison to traditional methods.

1. はじめに

近年の電子メールの普及とともに，大量の受信者に対して無差別的かつ一方的に電子メールを送信する，いわゆる迷惑メール（spam e-mail）が増加している．迷惑メールの増加率は電子メール全体の増加率を大きく上回っており，2001年には電子メール全体の10%以下だった迷惑メールが，2003年後半には全体の50%を超えたとの調査結果[2]も存在する．日本の電子メール利用者にとっても迷惑メール問題は他人事ではなく，日本においても迷惑メールに対する法的規制が行われているが，十分な効果をあげているとはいい難い．そのため，電子メールの利便性を保つうえで，迷惑メールを排除するための技術的対策が必要とされている．

迷惑メールに対する技術的対策として近年利用が増えているものの1つに，ベイジアンフィルタリングがある．これは電子メールをその内容から迷惑メールとそれ以外の正当な電子メールに分類する統計的フィルタリングの一種で，過去の電子メールからヘッダや本文などに含まれる単語などをトークンとして抽出し，各トークンの出現確率はすべて独立であるという仮定を設けたうえでベイズの定理を使用して各トークンに迷惑メール確率を設定し，各トークンの出現という事象（結果）からその電子メールが迷惑メールである，または正当な電子メールであるという原因を推定する手法である．

† 九州大学大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
†† 九州大学大学院システム情報科学研究院
Faculty of Information Science and Electrical Engineering, Kyushu University
☆ 現在，三菱電機情報ネットワーク株式会社
Presently with Mitsubishi Electric Information Network Corporation
☆☆ 現在，岡山大学大学院自然科学研究科
Presently with Graduate School of Natural Science and Technology, Okayama University

ベイジアンフィルタリングを用いた手法は文献 3), 4) など以前から研究が行われていたが，Graham の文献 1)，5) をきっかけにして，ベイジアンフィルタリングの実装が数多く開発されるようになった．これらの実装の中には，メールサーバに電子メールが到着した際に procmail [6] などを通じて実行されるものや，メーラがメールサーバにアクセスする際にプロキシサーバとして動作するもの，また既存のメーラの一部として動作するものなどが存在する．過去の電子メールに現れた単語と正当な電子メールにおける出現回数および迷惑メールにおける出現回数，学習した正当な電子メールの数および迷惑メールの数はコーパスと呼ばれるデータベースに格納される．

送受信される電子メールに含まれる言語が複数存在するような電子メール環境においてベイジアンフィルタリングを使用する場合，複数言語の存在を考慮せずに使用すると，電子メール中の特徴的なトークンではなく，電子メールに使用されている言語自体が分類の基準とされる恐れがあるため，日本語の電子メールを考慮した実装の多くは日本語とそれ以外でコーパスを分け，判定や学習の際にはまずその電子メールが日本語の電子メールか否かを調べ，その電子メールに対して使用するコーパスを選択する．一方，電子メール中には言語に依存しない正当な電子メールや迷惑メールの特徴が多く存在することが知られているが，上記のコーパスの分離を行った場合，それらの特徴は日本語の電子メールとそれ以外とで独立に学習や確率計算が行われる．一般にベイジアンフィルタリングは学習データが増すに従って精度が上昇すると考えられており，言語に依存しない特徴を日本語の電子メールとで合わせて学習・確率計算を行うことは精度向上に効果があると考えられる．

そこで本論文では，統計的フィルタリングの 1 つであるベイジアンフィルタリングに関して，判定精度をより向上させることができる手法を提案する．具体的には，日本語の電子メールと非日本語の電子メールが混在する環境において，言語ごとにコーパスを分離し，判定対象のトークンが属する言語のコーパスを用いて判定することにより，判定精度を向上させる．

2. ベイジアンフィルタリング

ベイジアンフィルタリングにおいては，まず，各トークンの迷惑メール確率を計算し，それらの確率をもとに判定対象となる電子メールの迷惑メール確率を計算する．そして，電子メールの迷惑メール確率が閾値を上回った場合に迷惑メールと分類し，閾値を下回った場合には正当な電子メールと分類する．Spambayes [7] などのように，閾値を 2 つ設定し，電子メールの迷惑メール確率が両方の閾値を上回った場合に迷惑メール，下回った場合に正当な電子メールと分類し，電子メールの迷惑メール確率が 2 つの閾値の中間にある場合には不確実と分類することで，正当な電子メールを失う危険性の低減を図った実装も存在する．

多くの実装では迷惑メール確率の計算の方法として Graham が文献 1)，5) で用いた方式や Robinson が提案した方式[8] が多く用いられている．これらの計算方式においては，トークン，および電子メールに対する迷惑メール確率は通常 0 から 1 の間の値をとるよう計算される．確率が 0 に近い値はそのトークンが正当な電子メールに特徴的なトークンであることを表し，その電子メールは正当な電子メールの可能性が高いということを意味する．確率が 1 に近い値は迷惑メールに特徴的なトークンであることを表し，迷惑メールの可能性が高いということを意味する．

迷惑メール確率の計算においては，まず，電子メール中に出現した各トークンに対する迷惑メール確率の計算を行う．この計算においては単純に正当な電子メール，迷惑メール内でのトークンの出現回数を比較するのではなく，コーパスに学習された正当な電子メール，迷惑メールの数を考慮しており，1 通の正当な電子メールにそのトークンが出現する確率と，1 通の迷惑メールにそのトークンが出現する確率を比較する形になっている．これは，コーパスに学習される正当な電子メールの数と迷惑メールの数の間で偏りが生じても計算される確率に偏りが生じないことを意図している．

電子メールに対する迷惑メール確率は各トークンに対する迷惑メール確率から計算される．上記の各方式では，高い迷惑メール確率を持つトークンを多く含む電子メールは高い迷惑メール確率を得る傾向があるほか，0 または 1 に近い迷惑メール確率を持つトークンは 0.5 に近い迷惑メール確率を持つトークンより重視され，また同一の電子メール内に複数回出現したトークンは 1 回しか出現しなかったトークンより重視されるように設計されている．

たとえば，Robinson の方式では，

$$p(w) = \frac{b/n_{bad}}{g/n_{good} + b/n_{bad}}$$

$$f(w) = \frac{s \times x + n \times p(w)}{s + n}$$

- x：$p(w)$ の平均
- n：正当な電子メールおよび迷惑メールの集合の

うち，単語 w を 1 つ以上含むメールの数

- s：適当な定数

を計算し，$f(w)$ をトークンの迷惑メールとする．次に，電子メールの迷惑メール確率 S を以下の式で計算する．

$$P = 1 - (1 - f(w_1)) \cdot (1 - f(w_2)) \cdots \cdot (1 - f(w_n))^{\frac{1}{n}}$$
$$Q = 1 - (f(w_1) \cdot f(w_2) \cdots f(w_n))^{\frac{1}{n}}$$
$$S = (P - Q)/(P + Q)$$

また，Robinson の方式の変形として，

$$H = 1 - C^{-1}(-2ln(\prod f(w)), 2n)$$
$$S = 1 - C^{-1}(-2ln(\prod (1 - f(w))), 2n)$$
$$I = (1 + H - S)/2$$

(C^{-1}: Inverse chi-square function)

と計算し，I を電子メールの迷惑メール確率とする方式があり，Robinson-Fisher 方式[8] と呼ばれている．

3. コーパスの選択手法の改良

3.1 コーパスの分離

複数の言語を取り扱う電子メール環境で統計的フィルタリングを行う場合，言語ごとにコーパスを分離しないと特定の言語で書かれた電子メールにおける迷惑メールの分類精度が低下し，統計的フィルタリングを使用するうえで許容できない割合の誤検出（正当な電子メールを迷惑メールに分類すること）や見逃し（迷惑メールを正当な電子メールに分類すること）が発生する恐れがある．これは，ある言語の電子メールにおける正当な電子メールと迷惑メールの比率（以後，迷惑メール含有率と呼ぶ）が，他の言語の電子メールにおける迷惑メール含有率と大きく異なる場合に発生する．たとえば，日本語の電子メールは主に正当な電子メールであり，英語の電子メールは多くが迷惑メールであるような環境では，学習される迷惑メールのほとんどが英語の電子メールとなる．このため，一般的な電子メールに使用される英単語などが迷惑メールに含まれていると，英語のメールはほとんどが迷惑メールのため，これらの単語の迷惑メール確率が高くなる．このことにより，一般的な電子メールに使用される英単語を含む正当な電子メールがことごとく迷惑メールと判定される事態が生じる恐れがある．

この問題に対処するために，従来の実装では言語によってコーパスを分離することが行われている．日本で開発された多くの実装[9),10)] では，電子メールを日本語の電子メールとそれ以外の言語の電子メールに分け，日本語の電子メールに対する確率計算・学習とそれ以外の言語の電子メールに対する確率計算・学習を独立に行っている．以後，この手法をメール単位でのコーパスの分離と呼ぶ．ここで日本語の電子メールとは，電子メール内に日本語の文字コードに属する文字が存在するものや，電子メールの表題や本文が日本語向けのエンコードが行われているものを指し，非日本語の電子メールは，これらの文字やエンコードが存在しないものを指す．つまり，日本語の電子メールには日本語以外の文字列も存在するが，非日本語の電子メールには日本語の文字列は含まれていない．このため，本研究では，日本語の電子メールに含まれる非日本語のトークンの扱いに着目した．

3.2 言語に依存しない特徴

迷惑メールのフィルタリングにおいて，電子メールのヘッダ部分にはフィルタリングに有用な情報（すなわち，正当な電子メールや迷惑メールに特徴的なトークンなど）が多く含まれていることが一般に知られている．電子メールのヘッダの仕様はインターネット上で共通であるため，この情報のうち多くはメールで使用されている言語に依存しない特徴であると考えられる．メール単位でのコーパスの分離を行った場合，この特徴もまた言語ごとに確率計算や学習が行われる．しかし，統計的手法を用いたフィルタリングにおいては，一般により多くの電子メールから学習データを作成することにより，より適切な迷惑メール確率が得られやすくなり，フィルタリングの精度を上昇させることができると考えられている．

そこで，本論文ではトークンごとにそのトークンの属する言語を判定し，その言語のコーパスを用いて確率計算・学習を行う方式を提案する．この方式は，ある電子メールに含まれるトークンについて，その電子メールが属する言語をもとにコーパスを選択するのではなくそれぞれのトークンについて個別に属する言語を判定し，日本語の電子メールと非日本語の電子メールに共通して現れるトークンに対する迷惑メール確率の計算を統一的に行うことによって，より多くの電子メールをもとに各トークンの迷惑メール確率の計算を行い，より高い精度を得ることを目的とした方式である．たとえば，日本語の電子メールの中に非日本語のトークンが現れた場合に，従来の実装では日本語のコーパスを用いてそのトークンの迷惑メール確率を計算し，学習の際も日本語のコーパスに反映していたが，提案方式では非日本語のコーパスを用いて判定・学習を行う．この方法は，コーパスを分けない場合に生じる前述の問題に対処しつつ，迷惑メールや正当な電子

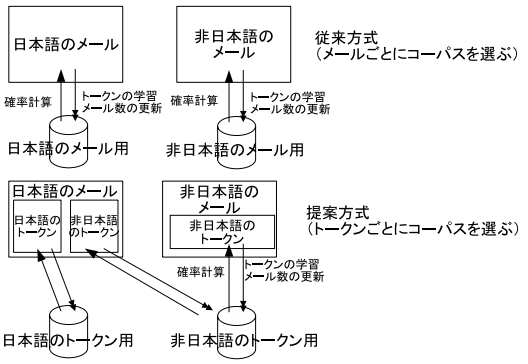

図 1 従来方式と提案方式のコーパスの使い方
Fig. 1 Usage of corpus in the traditional method and our proposed method.

メール中に存在する，言語に依存しない特徴（トークンの出現）をより効果的にフィルタリングに利用することを意図した方式である．従来方式と提案方式をは**図 1** のように図示することができる．

3.3 メール数の数え方について

提案手法を用いる際には，学習した電子メールの数をどのように数えるかという問題を考慮する必要がある．従来の手法では日本語の電子メールにおける確率計算や学習は，非日本語の電子メールにおけるそれと完全に独立であったので，日本語の電子メールを処理する際には現れたトークンを日本語のコーパスに追加したのに対して，提案方式では日本語の電子メールを処理する際には日本語のコーパス，非日本語のコーパスの両方に対してトークンを追加するためである．学習した正当な電子メールの数，迷惑メールの数を記録することはベイジアンフィルタリングにおいて非常に重要な役割を果たすため，この点をうまく処理することが精度の向上に結び付くと考えられる．

本研究では，直感的に考えられるいくつかの手法を用いてシミュレーションを行い，従来方式を含めて比較することにより，効率の良い電子メール数の判定方法を探索する．

提案方式では，日本語メールにおいてトークンごとに言語を判定し，登録するコーパスを選択する．このため，言語の割合を考慮せずそれぞれのコーパスに 1 通のメールとしてカウントする方法（方式 0）と，メールにおける各言語の割合に応じてメール数を数える方法が考えられる．日本語メールにおいては，日本語と非日本語に分けて学習させるため，メールに含まれる日本語と非日本語の割合を考慮し，各言語のコーパスにトークンを学習させることで，言語の出現割合に従って各トークンの迷惑メール確率を計算でき，精度を向上させることができると考えられる．このことを検証するため，メールにおける各言語の割合に応じてメール数をカウントする場合，メールから抽出した日本語のトークンと非日本語のトークンの数の比を加算する方法について評価した．たとえば，ある電子メールが 50 個の日本語トークンと 150 個の非日本語トークンを含むならば，その電子メールを学習する際には日本語のコーパスの「学習した電子メールの数」に 0.25 を，非日本語のコーパスの「学習した電子メールの数」に 0.75 を加算する．これを集計方式 1 とする．

表 1 実験を行った学習メール数の数え方
Table 1 Methods of counting messages.

	トークン数		メール数	
方式	日本語	非日本語	日本語	非日本語
方式 0	x	y	1	1
方式 1	x	y	x/(x+y)	y/(x+y)
方式 2	x	y	$\sqrt{x/(x+y)}$	$\sqrt{y/(x+y)}$
方式 3	x	y	$(x/(x+y))^2$	$(y/(x+y))^2$

しかし，実験に用いた環境では，4.1 節で説明する bigram を用いてトークンを抽出する．このため，トークン数がそのメールにおける言語の使用割合を表しているとはいえない．また，日本語と非日本語のトークン数の比が，各言語のそのメールにおける迷惑メールか否かの判断の寄与率を正確に表すとは限らない．したがって，日本語部分と非日本語部分の寄与率を求めるには，ベイジアンフィルタリング方式でそうであったように実験（経験）的にメール数を数える方法を評価し，求めることとした．メール数の数え方のヒントとして，トークン数の比を利用し，以下に示す 2 つの方法でトークン数の比を補正した手法を用い，提案手法において判定精度を向上させるメール数の数え方であるか否かを評価した．補正方法として，トークン数の比を計算し，その平方根をそれぞれ加算する方法，2 乗を加算する方法をそれぞれ試みた．これらを順に集計方式 2，集計方式 3 とする．集計方式 2 は各言語のメール数の値は大きくなるものの，日本語と非日本語の比は小さくなる方向に補正する．一方，集計方式 3 は各言語のメール数の値を小さくするものの，日本語と非日本語の比を大きくする方向に補正する．

表 1 はこれらの方式を比較したものである．

4. 実　　験

4.1 実験内容

以下のような条件で実験を行った．

実験に用いた電子メールは，筆者らが日常研究活動に用いているメールアドレスに受信した正当な電子

メールおよび迷惑メール，および筆者らのうち 1 名が所有するハニーポットアドレスで受信した迷惑メールである．その量は**表 2** のとおりであり，非日本語の電子メールとして英語の電子メールのみを使用している．日本語のトークンと英語のトークンは文字コードが重複しないので，個々のトークンごとにそのトークンの属する言語を判定することができる．これらの電子メールは複数メールアドレスから収集したため，Received，To など一部のヘッダについて整形を行い，収集元の情報がフィルタリングに用いられることのないようにした．

ベイジアンフィルタリングの実装として bsfilter[9]（Revision 1.35.4.13）を使用し，提案方式に応じて修正を加えて使用した．迷惑メール確率の計算式には Robinson-Fisher 方式（$s = 0.001$）を使用し，日本語からのトークンの抽出には bsfilter 内蔵の bigram を使用した．この bigram は以下のような規則でトークンを抽出する．

- 孤立した漢字および 2 字が連続する漢字，連続するカタカナはそのまま 1 つのトークンとして抽出する．
- 3 字以上の連続する漢字については 1 文字目と 2 文字目，2 文字目と 3 文字目というように隣接する 2 字の漢字をそれぞれ 1 つのトークンとして抽出する．
- 英単語については空白などで区切られた 1 単語を 1 つのトークンとして抽出する．

3.3 節で述べた 4 つの集計方式に加え，提案方式との比較対象として，従来のメール単位でのコーパス分離方式（Traditional method）とコーパスを分離しない方式（Single Corpus）について実験した．

実験の手順は以下のとおりである．

(1) 一定量の電子メールを，正当な電子メール・迷惑メールを明示してベイジアンフィルタに学習させる．この際，学習させる正当な電子メールと迷惑メールは用意したメールから毎回ランダムに選択し，その比率は用意した正当な電子メールと迷惑メールの数（表 2）に比例するものとする（初期学習）．

(2) 残りの電子メールを 1 通ずつベイジアンフィルタに判定させ，計算された電子メールの迷惑メール確率を記録する．

(3) (2) で記録した迷惑メール確率を集計し，判定対象となる電子メールが迷惑メールであると判定する閾値を 0 から 1 まで変化させて誤検出，見逃しの割合を測定する．

(4) 実験データの偏りによる影響を減らすため，上記の実験を 15 回試行し，測定値の平均を実験結果とする．

フィルタリングの精度はコーパスの学習量，すなわち，判定に使用する過去の電子メールの量を増すことで向上すると考えられている．前述のとおり，提案方式は非日本語のトークンについて日本語電子メール中での出現と非日本語電子メール中での出現を共通に取り扱うことによって，非日本語のトークンに対して迷惑メール確率を計算する際の「過去の電子メール」に相当する電子メールの量を増加させ，迷惑メールフィルタリングの精度を向上させることを意図した方式である．そこで，本実験では初期学習の量が異なる（2,000 通，500 通）2 通りの条件において実験を行った．

4.2 実験結果

従来方式（Traditional method，Single Corpus），提案方式（集計方式 0〜3）のそれぞれについて，横軸に誤検出の割合，縦軸に見逃しの割合をとり，閾値を変化させたときの両者の関係を示すグラフを**図 2**，**図 3** に示す．図 2 は正当な電子メールと迷惑メールから計 2,000 通を学習させたコーパスを用いた場合，図 3 は計 500 通を学習させたコーパスである．

また，従来方式と提案方式のそれぞれについて，日本語の電子メールと非日本語の電子メールを分けて誤検出の割合と見逃しの割合の関係を描くと**図 4**，**図 5**

表 2 実験に用いる電子メールの数
Table 2 Number of messages used in tests.

	日本語	非日本語
正当な電子メール	1,679	1,075
迷惑メール	267	994
合計	1,946	2,069

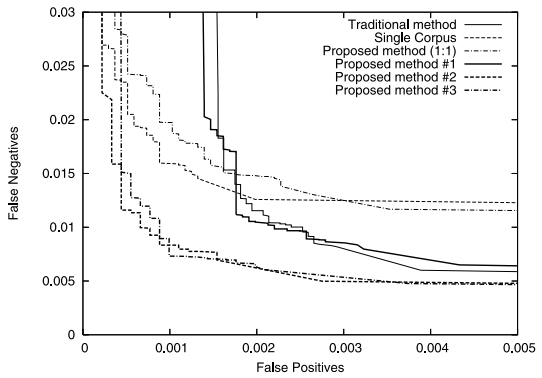

図 2 誤検出と見逃しの割合（学習量 2,000 通のコーパスを用いた場合）
Fig. 2 False Positives vs. False Negatives (initial learning: 2,000 messages).

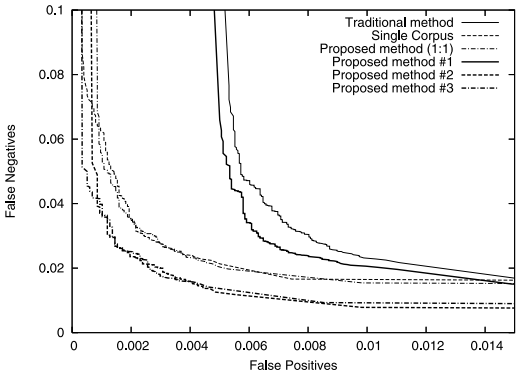

図 3 誤検出と見逃しの割合（学習量 500 通のコーパスを用いた場合）
Fig. 3 False Positives vs. False Negatives (initial learning: 500 messages).

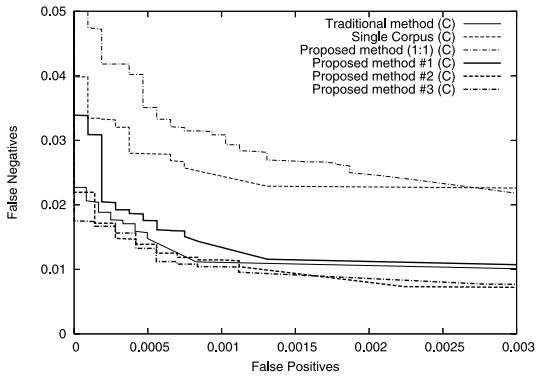

図 4 誤検出と見逃しの割合（学習量 2,000 通，非日本語）
Fig. 4 False Positives vs. False Negatives in non-Japanese (initial learning: 2,000 messages).

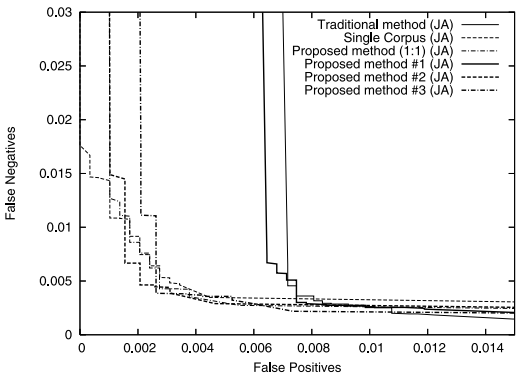

図 5 誤検出と見逃しの割合（学習量 2,000 通，日本語）
Fig. 5 False Positives vs. False Negatives in Japanese (initial learning: 2,000 messages).

のとおりになった．

メール単位のコーパス分離法（Traditional method）では，学習させたメール数に関係なく，日本語メールの判定において誤検出が特に多く，判定精度が悪い．これは，日本語メールに含まれる非日本語トークンに関しては，日本語メールについての学習データしかなく，非日本語トークンに対する学習データが不十分であったためだと考えられる．

すべての電子メールに対して同一のコーパスを用いた場合（Single Corpus），非日本語メールにおいて特に見逃しが多い．これは，3.1 節の冒頭で述べたように，実験に利用したメールは非日本語メールについては迷惑メールの割合が多いため，このように見逃しが増えたと考えられる．したがって，メールごともしくはトークンごとにコーパスを分離する手法が有効であることが分かる．

次に，提案手法を適用した場合について述べる．日本語と非日本語のトークンについて，それぞれを 1 通としてコーパスに学習させた場合（方式 0，Proposed method (1:1)），非日本語メールにおいて，見逃しが多くなっていることが分かる．日本語メールにおいて，非日本語のトークンの割合は比較的少ないと考えられる．このとき，日本語メールの非日本語のトークンにおいても 1 通と数え，学習させたため，非日本語トークンの寄与率が大きく計算されすぎ，判定精度が低下したと考えられる．

方式 1，2，3 には，いずれも非日本語メールに関して判定精度が良い．これは，日本語メールに含まれる非日本語トークンも学習しており，かつそのメール数の数え方もメールに含まれる割合に準じているからであると考えられる．

また，方式 2，3 に関してはいずれの場合も，日本語メールにおいても判定精度が良い．しかし，方式 1 においては従来方式の Traditional method よりわずかに誤検出確率が小さいものの，誤検出確率が比較的大きいといえる．各提案方式の差は，各トークンのメール数の数え方のみである．メール数の数え方においては，計算されるメール数の大きさの補正とメール数の比の補正が加わっており，方式 1 は日本語メールにおける非日本語トークンの寄与率計算が適切でなかったと考えられる．

以上のことから，以下のことが分かる．

(1) 提案方式（方式 1，2，3）はトークンごとにコーパスを選択して学習でき，非日本語トークンの学習量が増えていることもあり，非日本語メールにおける判定精度が従来方式とほぼ同程度の精度である．ただし，

日本語メールにおける非日本語トークンの数え方を言語によらず1通とした場合は，非日本語トークンの寄与率を大きく計算しすぎるため，精度が悪い．

(2) 提案方式は日本語メールにおいても判定精度が従来方式以上である．これは，日本語メールの非日本語トークンの迷惑メール確率計算に，非日本語メールの学習結果も利用できるためである．これにより，日本語メールの非日本語トークンの確率がより理想的な値に近づいていると推察できる．しかし，その判定精度はメール数の数え方の影響を受ける．

(3) 日本語と非日本語を合わせたメール判定精度においても，従来手法と同程度か，それよりも良い結果を出している．迷惑メール判定においては，正当なメールを迷惑メールと判定してしまう誤検出確率を減らすことが重要である．図2と図3から，提案手法（方式2，方式3）では見逃しの確率が従来手法と同程度でも，従来方式に比べて誤検出確率を4分の1程度に減らせることが分かった．したがって，提案手法を活用することにより，誤検出確率を大幅に減らすことができるため，本手法の有効性は高いといえる．

5. ま と め

本研究では，統計的フィルタリングの1つであるベイジアンフィルタリングに関して，複数の言語の電子メールを扱う環境における精度の向上について考察を行い，電子メール1通ごとではなく個々のトークンごとにコーパスを選択する方式を提案した．本方式の特徴は，従来一般的に行われていた，言語ごとにコーパスを完全に分けて確率計算や学習を行う方式と異なり，複数言語のトークンを含む電子メールについては個々のトークンの属する言語に応じてコーパスを使い分けることである．このことにより，電子メール全体に占める迷惑メールの比率が言語により異なる問題については従来どおりの対策を行いつつ，迷惑メールの特徴のうち言語に依存しないものを他の言語の電子メールに対する判定においても利用できることである．そして実験を通じて，学習した電子メール数の数え方（方式1）によっては，提案方式は従来方式に比べ，同等の性能を示す場合もあるものの，メール数の数え方（方式2，方式3）によっては，誤検出確率を従来方式の4分の1程度に減らすことができることを示した．また，日本語メールに含まれる非日本語トークン（ヘッダと本文中の非日本語）の学習データを非日本語メールの学習データと統合できるため，日本語メールにおける判定精度が向上することも示した．

本論文の実験において，提案手法の有効性を示したものの，メール数の数え方によっては判定精度が向上しない場合が見られた．このため，今後の課題として，メール数の数え方と判定精度の関係についてさらなる検討を行い，提案方式を使用するうえでより効果的な集計方式を明らかにする必要がある．

また迷惑メール対策に関する一般的な課題としては，迷惑メール送信者による迷惑メール対策回避のための手口への対策がある．ある迷惑メール対策手法が電子メール利用者の間で広く普及すると，その手法を回避するための手口もまた迷惑メール送信者によって広く利用されるようになると考えられる．実際，今回扱ったベイジアンフィルタリングをはじめとする統計的フィルタリングに対しても，すでに迷惑メール送信者によっていくつかの手口による回避が試みられていることが知られている[11]．統計的フィルタリング以外の迷惑メール対策手法を併用することを含め，新たなフィルタリング回避の手口の出現を監視し，継続的に迷惑メールのフィルタリングを有効に機能させる必要があるといえる．

謝辞 本研究の一部は，財団法人セコム科学技術振興財団平成15年度研究助成「インターネット妨害障害に対する暗号論的対策技術の研究」の支援を受けている．

参 考 文 献

1) Graham, P.: A Plan for Spam. http://paulgraham.com/spam.html
2) Brightmail: 50% of Internet E-Mail is Now Spam According to Anti-Spam Leader Brightmail. http://www.forrelease.com/D20030820/sfw036.P2.08202003020653.08038.html
3) Pantel, P. and Lin, D.: SpamCop—A Spam Classification & Organization Program, Learning for Text Categorization: Papers from AAAI Workshop, pp.55–62 (1998).
4) Sahami, M., Dumais, S., Heckerman, D., and Horvitz, E.: A Bayesian Approach to Filtering Junk E-Mail, Learning for Text Categorization: Papers from AAAI Workshop, pp.95–98 (1998).
5) Graham, P.: Better Bayesian Filtering, 2003 Spam conference. http://spamconference.org/proceedings2003.html
6) Procmail. http://www.procmail.org/
7) SpamBayes. http://spambayes.sourceforge.net/
8) Robinson, G.: Spam Detection. http://radio.weblogs.com/0101454/stories/2002/09/16/spamDetection.html

9) Bsfilter. http://www.h2.dion.ne.jp/˜nabeken/bsfilter/
10) Scbayes. http://namazu.org/˜satoru/scmail/scbayes.html
11) The Spammer's Compendium. http://www.jgc.org/tsc/

(平成 16 年 11 月 29 日受付)
(平成 17 年 6 月 9 日採録)

岩永　学

2003 年九州大学工学部電気情報工学科を飛び級のため中退．2005 年同大学大学院システム情報科学府情報工学専攻修了．同年三菱電機情報ネットワーク（株）入社．迷惑メール対策，ネットワークセキュリティの研究に従事．電子情報通信学会会員．

田端　利宏（正会員）

1998 年九州大学工学部情報工学科卒業．2000 年同大学大学院システム情報科学研究科修士課程修了．2002 年同大学院システム情報科学府博士後期課程修了．2001 年日本学術振興会特別研究員．2002 年九州大学大学院システム情報科学研究院助手．2005 年から岡山大学大学院自然科学研究科助教授．博士（工学）．オペレーティングシステム，コンピュータセキュリティに興味を持つ．電子情報通信学会，ACM 各会員．

櫻井　幸一（正会員）

1988 年九州大学大学院工学研究科応用物理専攻修士課程修了．同年三菱電機（株）入社．現在，九州大学大学院システム情報科学研究院情報工学部門教授．1997 年 9 月より 1 年間コロンビア大学計算機科学科客員研究員．2004 年 4 月より九州システム情報技術研究所第 2 研究室室長併任．暗号理論・情報セキュリティ・社会情報工学の研究に従事．博士（工学）．2000 年情報処理学会坂井特別記念賞，2000 年・2004 年情報処理学会論文賞，2005 年 IPA 賞受賞．電子情報通信学会，日本数学会，ACM，IEEE 各会員．