spamメールフィルタ

～ 日本語のspamメールに即したテキスト処理 ～

伊加田　恵志

近年，インターネットの発展と電子メールの普及に伴い，迷惑メール（一般にspamメールと呼ばれる）が増加している。電子メールは，非常に安価に大量に送信できることから，悪質な業者が受信者の望まない広告メールを不特定多数に送信しているためである。このようなメールが一方的に送信されることで，受信者は不要なメールを大量に受け取り，そのメールを削除するのに無駄な時間を費やしたり，また重要なメールがそれらのメールに埋もれてしまうために見逃したりという問題が起こっている。あるいは，子供が携帯電話を持つということがあたりまえになっているため，成人向けの内容を含むspamメールが届き，それが子供の目に入ってしまうという心配もある。

また，問題はメールを受信するユーザだけの問題にとどまらない。2001年にはインターネット上でやり取りされるメール全体の8%程度だったspamメールは，いまや全体の69%に達しているとの報告もある[1)]。このように大量に無駄なメールがネットワーク上に流れると，電子メールをユーザに届けるインターネット接続業者や携帯電話事業者などの電気通信事業者にとっても，設備の増設や運用体制の強化など余計なコストが必要となり，ユーザからの苦情への対応といった負担が増え問題となっている。このままでは，非常に便利なコミュニケーションツールである電子メールの仕組みそのものが破綻してしまう可能性もある。

現在，急増するspamメールに対しさまざまな対策が研究され，実用化されている。その中でも，初期から研究されているのが，自然言語処理技術を応用したフィルタリング技術である。

本論文では，立命館大学の協力を得て開発したフィルタリング技術[2)]について述べる。今回開発した技術は，日本のspamメールの状況をふまえ，その特徴を反映することで精度向上を図っている。

文書分類とspamメールフィルタリング

自然言語処理の分野において，「文書を自動的にあらかじめ与えられたカテゴリに分類」する文書分類技術が研究されてきた[3)]。これはたとえば，新聞記事を政治，経済，スポーツなどのカテゴリに自動的に振り分ける技術である。この技術を応用することで，spamメールフィルタリングを行うことができる（**図1**）。つまり，受信したメールの内容から，そのメールがspamというカテゴリに入るかど

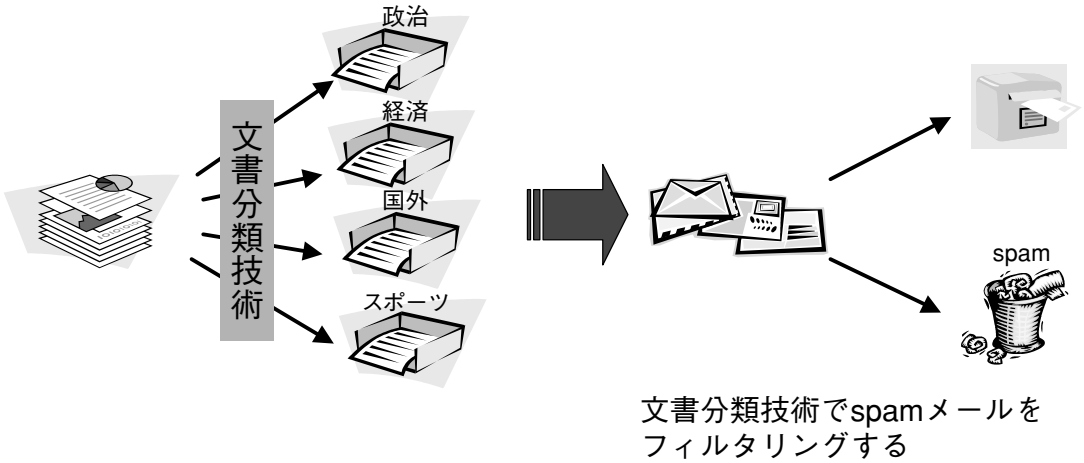

図1　文書分類技術とspamフィルタリング

うかを調べればよいこととなる。

spamフィルタリング技術の中でも特に広く知られているのが，Grahamによって提案されたベイジアンフィルタである[4]。この技術は，ベイズ理論を元にした機械学習型のフィルタである。そのほかにも，文書分類では，決定木や，サポートベクタマシン（SVM）と呼ばれるものが分類学習器として使われることが多い。特にSVMは，近年，自然言語処理などの分野で有効性が高い学習アルゴリズムとして注目されている。SVMは，学習用のデータを正例と負例の2つに分けてベクトル空間上に配置し，その正例と負例の2つの空間を，それぞれの事例のうち最も近いものを取り出して，その両事例の距離が最も大きくなるように境界線（面）を決める機構をもつ。spamメールを負例，正当なメールを正例とすることで，境界線を学習し，spamメールを判別する。今回開発したシステムでは，このSVMを利用した。

日本におけるspamメールの状況

わが国におけるspamメールの特徴として，携帯電話に対し送信されるものが非常に多いということがあげられる[5]。また，spamメールの内容としては，出会い系サイトやアダルト商品の販売などが多く占められているという点が特徴的で，受信者への不快感や，昨今の出会い系サイトを利用した犯罪の温床になると考えられる点も問題になっている（**図2**）。

そこで，日本語のspamメールの内容にどのような特徴があるのか分析した。分析には,参考文献2）に示された，2002年4月14日から2002年5月13日までに実際に携帯電話に送信されたメールを利用した。

分析の結果（**表1**），これらのspamメールには,WWW

図2　日本語spamメールの例

表1　メール分析結果

文字数	123文字/1通
URL出現数	2.1回/1通
メールアドレス出現数	0.1回/1通
伏せ字出現数	0.05語/1通

（World Wide Web）サイトのアドレス（URL）や電子メールアドレス，また，単語の一部を記号に置き換えた伏せ字が特徴的に現れ，それらがspamメールを判定するときの鍵になるのではと考えられた。これらの特徴からspamメールの前処理を行うことが，フィルタリング精度の向上に繋がるのではないかと推測した。

spamメールフィルタシステム概要

図3に，開発したspamメールフィルタシステムの概要を示す。システムは大きく以下の3つの部分から構成されている。

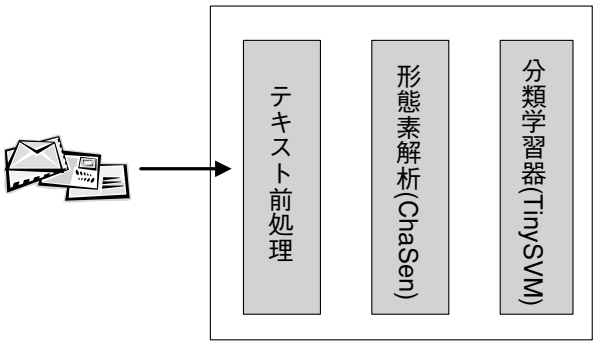

図3　spamメールフィルタシステム構成

テキスト前処理部
形態素解析部
分類学習器

メールテキスト前処理部は，前節で述べた日本語spamメールの特徴と思われるURLや電子メールアドレス，伏せ字を分類学習器の学習に取り入れ，既存のフィルタリング手法より判定精度を向上させる目的で開発した。詳細は次節で述べる。

形態素解析部は，入力された日本語メールテキストを，単語に区切る部分である。本システムでは，奈良先端科学技術大学院大学で開発されたChaSen[6] を利用した。

spamメール分類学習器は，あらかじめ用意したspamメールとそれ以外のメールを学習させることにより，spamメールに属するかどうかを判定する部分である。本

システムではサポートベクタマシンSVMの実装の一つであるTinySVM[7] を利用した。

メールテキスト前処理

テキスト前処理部では，spamメールに特徴的に現れる，URL， 電子メールアドレス，伏せ字などの表現を検出する。これらの文字列は，人間にとっては単語として認識することができるが，機械にとってはただの文字の連続にしか認識することができず，形態素解析を行うとただの英数字や記号としてばらばらに区切った結果を出力してしまう。このままでは，せっかくの特徴を学習に反映することができない。そこで，まず検出した表現を以下のような文字列に置換するようにした。

URL	→	mwkURL
電子メールアドレス	→	mwkMAIL
電話番号	→	mwkTEL
伏せ字	→	mwkFSJ

これらの置換した文字列を形態素解析器の辞書に単語として登録しておくことで,置換した文字列を一つの単語として出力することが可能となる。**図4**に前処理を行ったメール本文の例を示す。

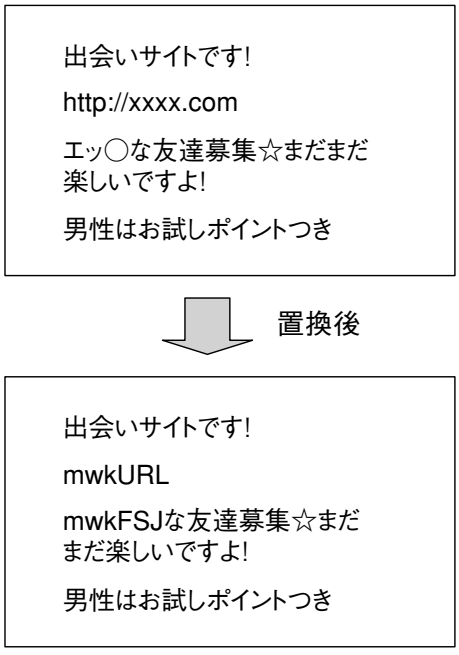

図4　置換例

実験および精度評価

開発したspamメールフィルタシステムに対し，評価実験を行い,判定精度の測定を行った。また，比較のため，URLなどの置換を行う前処理を行った場合と，行わなかった場合でそれぞれ実験を行った。実験に用いたデータは，

spamメール　6134通
正当なメール 3878通

である。これらのうち，spamメールの2879通と正当なメールの225通を学習データとして用い，残りの6908通（内訳：spamメール3255通，正当なメール3653通）を判定した。

評価は，再現率（*Recall*），適合率（*Precision*），精度（*Accuracy*）を用いて実施した。再現率はやってきたspamのうちspamと判定できた割合を表し，この値が大きいと，spamの取り逃しが少ないことを表す。また，適合率はspamと判定したうち本当にspamであるものの割合で，この値が大きいと正当なメールをspamと誤判定する確率が小さいことを表す。精度はspamとそうでないメールをどれだけ正しく振り分けられたかというフィルタの性能を表す。

$$Recall = \frac{N_{tp}}{N_{tp} + N_{fn}}$$

$$Precision = \frac{N_{tp}}{N_{tp} + N_{fp}}$$

$$Accuracy = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}}$$

ここで，N_{tp}，N_{tn}，N_{fp}，N_{fn}はそれぞれ，

N_{tp}：システムがspamメールをspamと判定した数
N_{tn}：システムが正当なメールを正当と判定した数
N_{fp}：システムが正当なメールをspamと判定した数
N_{fn}：システムがspamメールを正当と判定した数

を表す。

表2および，**表3**は，実験から得られたメールの判定結果である。

表2　メール判定結果

	N_{tp}	N_{fp}	N_{fn}	N_{tn}
前処理あり	2876	84	4	3944
前処理なし	2878	376	2	3652

表3　判定精度評価結果

	再現率	適合率	精度
前処理あり	99.9%	97.2%	98.7%
前処理なし	99.9%	88.4%	94.5%

前処理あり，前処理なしとも全体的に高い精度で判定ができていることがわかる。再現率において，前処理の有効性は確認することはできなかったものの，適合率では約9ポイントの違いが現れ，全体の精度として4ポイントの差になっている。前処理を行うことにより，spamメールの判定漏れが改善でき，実験では，より，376通見逃していたspamメールのうち，292通を正しくspamと判定できるようになったことがわかる。

ただ，前処理ありにすると,システムが正当なメールをspamと判定した数N_{fn}がわずかながら増えている。これは，ユーザにとって必要なメールを見逃すことにつながり，どのspamメールフィルタリング技術においてもできる限り抑えたい数である。今後，実験結果などを詳しく分析して，この数を抑えるように更なる工夫が必要と考える。

お わ り に

日本におけるspamメールの特徴を学習機構に取り入れたspamメールフィルタリング技術について，その機能および精度評価を述べた。この技術は，実験により，98.7%という高精度でspamメールを判別できるという結果が得られた。

spamメールも年々巧妙化しており，フィルタリングを簡単にすり抜けようと，文字列を工夫するなど回避策をほどこしている。フィルタリング以外にも，インターネット接続業者や携帯電話事業者において，同時送信数の抑制や，メールアドレス・ドメイン指定拒否などさまざまなspamメール対策を行っている。しかし，これらも当初は一定の効果を見せるものの，新たな回避策によって有効ではなくなってしまう。現在，このような対抗策とその回避方法との戦いはいたちごっこの様相を呈している。今後も，spamメールの動向を調査し，新たな技術開発を行うことが必要とともに，spamメール対策を推進する企業，団体，政府が協力して取り組んでいくことが課題となろう。 ◆◆

TIPS【基本用語解説】

形態素解析（けいたいそかいせき）
形態素解析とは，自然言語処理の基礎技術のひとつで，文を単語に分割し，品詞付けを行う処理のことである。特に，日本語は，英語のように単語同士が空白で区切られてないので，自然言語処理を行う上で非常に重要な技術となっている。

ベイズ理論
ベイズ理論とは，「過去に起きた事象の発生頻度（確率）から未来の出来事の発生頻度（確率）を予測する」というものである。

ベイジアンフィルタ
ベイズ理論を使った学習型spamフィルタの一種で，あらかじめ過去に受信したメールの内容からspamに属する確率を学習し，今受け取ったメールがspamに属する確率を予測（計算）するものである。

■参考文献

1）http://www.postini.com/stats/
2）増田 明宏，福本 淳一：“学習機構を用いた迷惑メールの分類”第3回情報科学技術フォーラム 一般講演論文集 第2分冊，電子情報通信学会，pp.215–217，2004年
3）永田 昌明，平 博順：“テキスト分類 – 学習理論の「見本市」–”，情報処理，Vol.42 No.1，情報処理学会，pp.32–37，2001年
4）Graham, p: “A plan for Spam”，
http://www.paulgraham.com/spam.html, 2002
5） David Crocker，景山 忠史，山井 成良，鈴木 常彦，安藤 一憲，中村 泰典，加藤 佳実，伊藤 孝史，近藤 学，岡村 久道: “spamメールの現状と対策の動向”，情報処理，Vol.46，No.7，pp.739-791，2005，情報処理学会
6）ChaSen：http://chasen.naist.jp/hiki/ChaSen/
7）TinySVM：http://chasen.org/~taku/software/TinySVM/

●筆者紹介

伊加田恵志：Satoshi Ikada. 研究開発本部 ユビキタスシステムラボラトリ