COLUMN

インターネット生活向上委員会

01

SPAMを防ぐ

山本 文治（㈱IIJメディアコミュニケーションズ）
bunji@iij-mc.co.jp

今回から「インターネット生活向上委員会」として，連載を始めることとなりました．ますますコモディティとなってゆくインターネットですが，発展途上のインフラストラクチャであるがゆえに，まだまだ不便なことも多いです．この連載では「インターネットのある暮らし」の生活品質を高めていくことを目的とし，さまざまなアプリケーションなどを紹介していきます．

◉迷惑メールの問題点◉

インターネットや携帯電話に代表される電子メディアの中で，もっとも活用されるアプリケーションは電子メールだろう．携帯電話を使って熱心に電子メールをやりとりする姿は日常のものとなり，電子メールアドレスが刷り込まれていない名刺も見かけなくなった．

このように便利なアプリケーションにも近年1つの問題が浮上してきた．「迷惑メール」（SPAMとも呼ばれる）の存在である．すでに社会問題化しており，2002年には「特定電子メールの送信の適正化等に関する法律」が施行され，同意・承諾のない受信者に対して広告・宣伝メールを送る際のフォーマットが規定された．また「特定商取引に関する法律」の一部においても広告メールに関しての改正がなされるなど，法制的な取り組みも進んでいる．

迷惑メールの送付は国際的な広がりを見せており，海外から送付されてくることも多い．迷惑メールを発信する事業者はさまざまな方法でメールアドレスを収集し，送付してくる．でたらめなメールアドレスをランダムに生成し送りつけるという乱暴な手法も行われており，迷惑メールの数は増える一方だ．迷惑メールを排除するために多大な時間を割かれるようになり，電子メールの整理はいまや悩ましい作業となってきているほどだ．

◉迷惑メールを防ぐ～Bayesianフィルタ◉

迷惑メールを防ぐため，サーバ側とクライアント側の双方で対策がされてきた．最も初歩的なものは次の2つだろう．

- 許可されていない第三者へ／からの中継を許可しない（サーバ）
- 送信者についてホワイトリスト／ブラックリストを作成（クライアント）

だがこれらの手法だけでは，押し寄せる迷惑メールは防ぎきれない．迷惑メールの発信元となるメールサーバは世界中に分散しており規制しきれない．リストのアップデートを続けていくのも大変な労力が必要だ．

そこで，最近注目を集めているのがBayes理論を元にした「Bayesianフィルタ」だ．Bayes理論は過去の統計を元にして結果を予測する確率論であり，この手法を迷惑メールの判定に応用したものがBayesianフィルタだ．メールの1通1通について「迷惑メールに使われそうな語句がどれくらいあるか」を判定し，迷惑メールか否かを総合判断するのだ．受け取ったメールのうち迷惑メールとそうでないメールの語句から，アップトゥデートな辞書を作成していく．人によって受け取る迷惑メールの量や傾向は異なるため，あくまで本人が受け取ったものから作成する方がよいといわれている．このように辞書を運用していくうちにヒット率が高まっていくというわけだ．

◉POPFileを使ってみる◉

実際にWindows上で稼働するBayesianフィルタであるPOPFileを導入してみる．POPFileは以下の特徴を持つ．

1. クライアントと同じマシンで稼働し，クライアントとPOPサーバ間でプロキシとして動作する．
2. 迷惑メール振り分けだけではなく，メールボックス自動振り分けツールとしても使用できる．
3. 日本語メールに対応している．
4. Webブラウザを経由して設定できる（日本語表示）．
5. POP3に対応しているメールリーダならばどれでも組み合わせられる．
6. Perl版も用意され，広範なプラットフォームにも対応．

POPFileは「受信時にメールの中身をスキャンし，それが迷惑メールと判定できたらメールにマークを付け，メールをクライアントに渡す」という作業を行う．

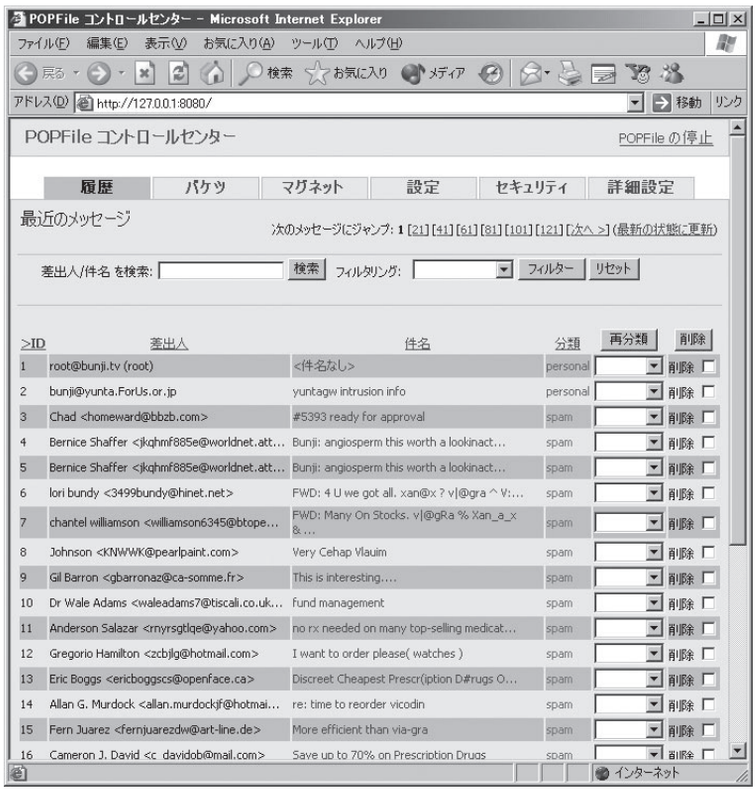

図 -1

Bayesianフィルタは学習が必要だが，POPFileの場合スキャンしたメールについて人間が分類するという作業をとる．

インストーラはhttp://popfile.sourceforge.net/ より入手できる．POPFileの最新バージョンは 0.20.1 だが，バグがあるため日本語化対応パッチもhttp://popfile.sourceforge.jp/ より入手し，インストールする．POPFileは自分のPC上でプロキシサーバとして稼働するため，クライアントにてメールサーバの設定を変更する．POPFileは「迷惑メールか否か」という判定を実施すると，結果表示としてそのメールの「X-Text-Classification」ヘッダ，ないし「Subject」ヘッダに「[spam]」というようなタグを挿入する．クライアントの自動振り分け機能と連携させれば，迷惑メールを特定のフォルダに自動的に分別することができるようになる．

◉POPFileの運用◉

POPFileには届いたメールを人間が迷惑メールかどうかをチェックするインタフェースが備えられている．しばらくの間効果は上がらないかもしれないが，これを続けることにより迷惑メールの判定率が上がるだろう．

筆者はこの原稿のためにPOPFileを 2 カ所（職場と個人の 2 つのメールアドレス）で導入し，運用してみた．するとどちらも導入後 1 週間ほどで迷惑メールの判定率が 95% 近くまで上がった．受け取るだけでも不愉快な迷惑メールだが，自動的に排除できる仕組みがあれば仕事の能率も上がる．

だが 5% の誤判定があるのも確かだ．迷惑メールをそうでないと判定されるのはまだしも，真っ当なメールを迷惑メールと判定され，読み飛ばしてしまうとダメージが大きい．こうしたBayesianフィルタを実環境に導入する場合は，迷惑メールと判定したメールをファイルするフォルダをゴミ箱などとは別に作成しておき，時間のあるときに漏れがないかチェックすることを奨めたい．筆者のまわりではこうしてメールの読み落としを防いでいる人が多かった．

◉他の迷惑メール対策◉

Bayesianフィルタを内蔵しているメールクライアントも増えてきた．EudoraやThunderbirdなどだ．またBayesianフィルタだけではなく，あらかじめ迷惑メールやウィルスメールの特徴的なデータを保有し，排除できるフィルタやクライアントもある．深刻化しつつある迷惑メール問題だが，対策ソフトウェア側でもBayesianフィルタだけでなく，複数のアプローチが用意されていくだろう．

Bayesianフィルタのアプローチは，迷惑メールの判定以外にも応用が効くと考えられている．検索エンジンやパーソナライジングという分野だ．いずれにせよ，我々の生産性，生活の向上に役立ってくれそうなテクノロジーである．

（平成 16 年 3 月 1 日受付）