

ベイジアン方式と機械学習を併用した スパムメールフィルタの検討

山口 博之^{†1} 角 朝香^{†1,*1}
杉井 学^{†2} 松野 浩嗣^{†1,†2}

近年の迷惑メールの増大は、世界中の電子メール利用者を悩ませている。スパムメール対策として様々なメールフィルタの開発が行われているが、その中で現在最も広く利用されているのが、ベイズ理論に基づいてスパムメールを分類するベイジアンフィルタである。bsfilter は、ベイジアンフィルタの一つであるが、多くのプラットフォームに対応できることなどから、利用者が増えている。しかし bsfilter は、正規メールの正解率が高いが、スパムメールの正解率が低い特徴がある。そこで、先に我々が開発した機械学習を応用したメールフィルタリングと bsfilter を組み合わせ、さらに bsfilter のメール判定に用いられる閾値の変更を行うことで正規メールとスパムメール共に、正解率の高いメールフィルタリング手法を検討した。

Investigation of Spam Mail Filtering Systems with the Use of Bayesian System and Machine Learning System

HIROYUKI YAMAGUCHI,^{†1} SAYAKA KADO,^{†1,*1}
MANABU SUGII^{†2} and HIROSHI MATSUNO^{†1,†2}

Recently, increasing of spam mails leads us to serious problems all over the world. Various mail filters had been developed for spam mails, among them a bayesian filter based on the bayes theory are generally used in these days. Bsfilter is one of the bayesian filters and it has been very popular system because it can be applied to many platforms. Bsfilter can classify necessary mails with high accuracy rate but it can not classify spam mails with such a high accuracy rate. In this study, we discussed a method for mail filtering system that combines bsfilter and machine learning system. Then, we investigated the combination of these two systems and alteration of a threshold for classification by the bsfilter in order to develop a new filtering system which can classify necessary mails and spam mails with high accuracy rate.

1. はじめに

電子メールはインターネットや携帯電話の普及により多くの人に利用されている。しかし、受信者にとって必要でないメール、いわゆるスパムメールが電子メール利用者を悩ませている。スパムメールは電子メールの特徴である一度に大量に送信できる点などを悪用して無差別に送りつけられるが¹⁾、その問題点は、受信者にとって必要でないメールが大量に送られてくるのが原因で、必要なメールを見落とす危険性などにある。スパムメールの送信手法は様々存在するが、送信者情報を偽ったメールや知人からのメールを装ったケースなどが増えてきており、手口は年々巧妙化してきている²⁾。

スパムメールに対抗する手段として、スパムメールと必要な正規メールを自動的に分類する様々なスパムメールフィルタリングが開発されている。その中で現在最も広く利用されているのが、ベイズの定理に基づいてスパムメールを分類するベイジアンフィルタである。bsfilter³⁾ はベイジアンフィルタの一つで正規メールの回収率は非常に高いが、スパムメールの正解率が低い。なぜなら、正規メールの誤判定を可能な限り少なくするために、多くの判定の難しいスパムメールを正規メールへと分類しているからである。現在、スパムメールの流通量は正規メールのそれよりはるかに多く、スパムメールの誤判定による正規メールへの混入をより少なくすることは、メール利用者にとって非常に重要である。

スパムメールの誤判定を少なくするためには bsfilter の分類判定に用いられる閾値を変更し、通常正規メールと誤判定されるメールをスパムメールと判定されるように変更すればよい。しかし、それだけでは正しく判定できていた正規メールを誤判定する可能性があるため、bsfilter と先に我々が開発した機械学習を応用したメールフィルタリング⁴⁾ を組み合わせることで、正規メールの正解率を維持したまま、スパムメールの正解率を上昇させる手法について検討する。

2. 機械学習システム BONSAI

機械学習システム BONSAI は、概念学習の一つである確率的近似学習 (Probably Approximately Correct Learning : PAC 学習) の学習パラダイムに基づき開発されたシステ

^{†1} 山口大学大学院 理工学研究科 Graduate School of Science and Engineering, Yamaguchi University

*1 現在、日本ラッド株式会社 Presently with Nippon RAD Inc.

^{†2} 山口大学 大学情報機構 メディア基盤センター

Media and Information Technology Center, Yamaguchi University

ムである⁵⁾。BONSAIは、一次元の記号列データを対象としており、あらかじめ分類しておいた正の例と負の例と呼ばれる二つの学習例を与えることで、それらを分類する規則を導き出す。

BONSAIは「インデキシング」と「決定木」により分類規則を作成する。インデキシングは学習例に含まれる要素に対してグルーピングを行い、要素数を減少させる。学習例の特徴を維持したまま要素数の減少ができれば、学習の効率化につながる。また、決定木はインデキシング後の記号列をノードに配置し、正の例と負の例を最も高い確率で分類することができる規則を提示する。

3. BONSAIを応用したメールフィルタリング

我々はこれまでに、BONSAIとRough Reading⁶⁾の手法を用い、スパムメール群と正規メール群を学習例としてBONSAIに入力し、二つのメール群を分類する特徴パターンの抽出を試みている⁴⁾。与えられた文章中の各単語は、以下の方法で出現数や出現頻度を基に重要単語の選定を行い、BONSAIに入力される。以下では簡単に、BONSAIを応用したメールフィルタリングの処理の流れについて説明する。

3.1 単語の抽出

学習用メールとして分類したスパムメール群、正規メール群に含まれる単語の抽出を行う。スパムメール群（正規メール群）の抽出した重複を含むすべての単語数を $E_s(E_h)$ とし、スパムメール群（正規メール群）の抽出した単語 w_i の出現数を $C_s(w_i)(C_h(w_i))$ とする。なお、スパムメール群と正規メール群ではメール中から抽出する単語数が大きく異なる場合が多い。そのため式(1)を用いて、スパムメール群と正規メール群の抽出単語数を基に補正を行い、スパムメール群、正規メール群での抽出単語 w_i の出現数を再計算する。スパムメール群の抽出単語 w_i の出現数 $N_s(w_i)$ は式(2)、正規メール群の抽出単語 w_i の出現数 $N_h(w_i)$ は式(3)から算出する。また、学習メール群全体での抽出単語 w_i の出現数 $N_{all}(w_i)$ は式(4)から算出する。

$$r = \frac{E_s}{E_h} \quad (1)$$

$$N_s(w_i) = C_s(w_i) \quad (2)$$

$$N_h(w_i) = C_h(w_i) \times r \quad (3)$$

$$N_{all} = N_s(w_i) + N_h(w_i) \quad (4)$$

3.2 単語の出現率と出現頻度の算出

学習例のメール群に含まれる各単語の出現率とスパムメール群での出現頻度をそれぞれ算出する。単語 w_i の出現率 $rate(w_i)$ は式(5)、スパムメール群での単語 w_i の出現頻度 $p(w_i)$ は式(6)から算出する。

$$rate(w_i) = \log(N_{all}(w_i)) \quad (5)$$

$$p(w_i) = \frac{N_s(w_i)}{N_{all}(w_i)} \quad (6)$$

3.3 アルファベット変換

単語を記号列へと変換する場合、学習メール群全体に一定回数以上出現した単語について、スパムメール群での単語の出現頻度を基に変換を行う。前節で算出した出現率 $rate(w_i)$ が基準値 $rate_base$ 以上の値を持つ単語 w_i を記号列変換の対象単語にし、基準値 $rate_base$ は式(7)によって算出する。なお、学習例のメール群に含まれるすべての単語の中で、出現率の最大を $rate_max$ とする。

$$rate_base = \frac{rate_max}{2} - 1 \quad (7)$$

スパムメール群での単語の出現頻度 $p(w_i)$ の値を基に、表1のように「x, y, z, a, b」の5つの記号からなるアルファベットに置き換える。つまり、「x, y」に置き換えられた単語はスパムメール群に特徴的に現れる単語、「a, b」に置き換えられた単語は正規メール群に特徴的に現れる単語、「z」に置き換えられた単語はスパムメール群、正規メール群の両メール群に特徴的に表れる単語である。

表1 出現頻度による記号列変換表

Table 1 Conversion of word based on an appearance frequency of the word.

変換記号	$freq_s(word)$
x	0.8 以上
y	0.8 未満, 0.6 以上
z	0.6 未満, 0.4 以上
b	0.4 未満, 0.2 以上
a	0.2 未満

なお、「x, y, z, a, b」のいずれにも置き換えられなかった単語に対しては、「o」で置き換える。単語をアルファベットに変換した結果、学習例に含まれるそれぞれのメールは「x, y, a, b, z, o」の6つの記号からなるアルファベットで表現される。

3.4 BONSAI による分類規則の作成

3.3 節で作成した学習例のメール群を BONSAI に投入し、スパムメール群と正規メール群を分類する規則の生成を行う。BONSAI は学習例を構成する「x, y, z, a, b, o」の 6 つの要素に対して、更にインデキシングを行い、学習例であるスパムメール群と正規メール群を最も効率よく分類できる規則を決定木として出力する。

4. ベイジアンフィルタ

ベイジアンフィルタはスパムメール群と正規メール群を学習例として、それぞれのメール群に出現した各単語とその出現頻度から、それぞれの単語がスパムメールに出現する確率を算出する。判定対象とするメールに出現した各単語に対する単語のスパム確率をもとに、そのメールがスパムメールである確率を算出する。メールのスパム確率があらかじめ定めた閾値以上であればスパムメール、そうでなければ正規メールと判定する⁷⁾。

5. BONSAI と bsfilter による併用フィルタリング

BONSAI と bsfilter の併用を行い、スパムメールと正規メールを分類する併用フィルタリングについて説明する。2 つのフィルタはそれぞれが異なる判定方式を採用しているため、一方がメールを誤判定しても、もう片方が正しく判定できるのではないかと予想した。BONSAI を用いたフィルタリング手法は、学習群のメール中に含まれる単語の出現頻度と語順により分類規則を作成し、メール中に特定の記号列のパターンが含まれているかにより判定を行う。つまり、BONSAI は決定木のノードパターンにマッチするかによってのみ判定を行い、メールの強い特徴を表す一部分のみが判定に利用されている⁴⁾。一方、bsfilter は学習群のメール中に含まれる各単語の出現頻度によって、総合的にそのメールがどれぐらいスパムメールであるかの確率を算出している⁷⁾。つまり、bsfilter はメール全体の文字列情報から、メールの全体的な類似傾向を利用して判定を行っている。BONSAI はメールの局所的な情報を利用しているのに対して、bsfilter はメールの平均的な情報を利用しているため判定方式が異なっている。また、語順を判定に利用する BONSAI を併用することで、ベイジアンフィルタの回避策として使用されているワードサラダ⁸⁾にも対応できる可能性がある。

分類正解率については、BONSAI は正規メールおよびスパムメールいずれの正解率も高いが、少数の正規メールを誤判定する傾向にある。一方、bsfilter は正規メールにスパムメールを多く混入してしまう傾向にあるが、正規メールの回収率は、100%に近い。そこで、併用フィルタリングでは、二つの手法の長所を組み合わせ、弱点を補い合って分類制度の高いフィ

ルタリング手法を確立することを目的とする。図 1 に bsfilter と BONSAI の併用フィルタリングの流れを示す。ham は正規メール、spam はスパムメールを表す。

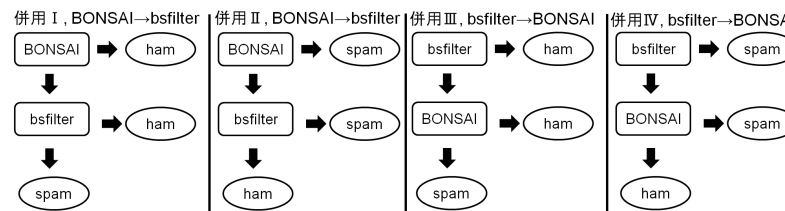


図 1 BONSAI と bsfilter による併用フィルタリングの流れ
Fig. 1 Process of Spam Mail Filtering System with BONSAI and bsfilter.

図 1 の併用 I は一段階目の BONSAI が正規メールと判定したメールに対してはその時点で判定を決定し、スパムメールと判定したメールに対しては二段階目の bsfilter が判定を行う。一段階目の BONSAI がスパムメールを正規メールと誤判定するメールは少ないが、二段目の bsfilter により、正規メールと誤判定するスパムメールが増えてしまう。併用 II は一段階目の BONSAI がスパムメールと判定したメールに対してはその時点で判定を決定し、正規メールと判定したメールに対しては二段階目の bsfilter が判定を行う。一段階目の BONSAI がスパムメールと判定したメールの中に正規メールが含まれ、これらのメールは回収することができないことが大きな欠点である。併用 III は一段階目の bsfilter が正規メールと判定したメールに対してはその時点で判定を決定し、スパムメールと判定したメールに対しては二段階目の BONSAI が判定を行う。bsfilter は正規メールを正しく判定するため 100%の正規メールを正しく回収することが期待できるが、正規メールと判定したメールの中に大量のスパムメールが含まれてしまうことが難点である。併用 IV は一段階目の bsfilter がスパムメールと判定したメールに対してはその時点で判定を決定し、正規メールと判定したメールに対しては二段階目の BONSAI で判定を行う。bsfilter はスパムメールを正しく分類できるが、正規メールの判定が BONSAI によることになり、少数のスパムメールと少数の正規メールを誤判定してしまう危険性が残る。

6. bsfilter の判定結果による閾値変更

本章では閾値変更前の bsfilter、BONSAI によるメールフィルタ、及びこれらの併用フィルタで判定を行った結果について考察する。また、併用フィルタによる問題点を検討し、メー

ルのスパム確率の分布により bsfilter の閾値を変更する方法を提案する。

6.1 使用メールサンプルデータ

実験には、2007 TREC Public Spam Corpus (TREC07)⁹⁾ の partial データセットを使用した。TREC07 の partial データセットは 2007 年 4 月 8 日から 2007 年 7 月 6 日のメール 30338 通 (スパムメール 6280 通, 正規メール 24058 通) から構成される。メールは時系列順に収められており、実験を行う際も時系列順に取り扱った。partial データセットの時系列の頭からスパムメール、正規メールそれぞれ 1000 通ずつ選出し、学習用メール群とした。さらに、時系列順にスパムメール 1000 通, 正規メール 1000 通を分類対象メール群として判定を行った。

6.2 bsfilter の閾値変更前の判定結果

学習数を 1000 通 (スパムメール 500 通, 正規メール 500 通) として判定を行った結果を表 2 に示す。学習メール数は表 1 中の「training mail」で表す。分類対象メール群を分類したシステム名を「System」で示す。bsfilter の後に書かれた数字値はメール判定に用いられる閾値の値である。BONSAI によるメールフィルタリングは BONSAI, bsfilter と BONSAI の併用フィルタリングは併用、また、併用の後に書かれている番号は図 1 の併用フィルタの番号と対応している。分類対象メール数, 分類対象スパムメール数, 分類対象正規メール数はそれぞれ「test_{total}」, 「test_{spam}」, 「test_{ham}」で表す。correct_{spam}(correct_{ham}) は「スパムメール (正規メール) を正しく分類したメール数」である。error_{spam}(error_{ham}) は「スパムメールを正規メール (正規メールをスパムメール) と誤って分類したメール数」である。AccuracyRate_{spam} は「スパムメールを正しく分類した割合」であり式 (8) から、AccuracyRate_{ham} は「正規メールを正しく分類した割合」であり、式 (9) から算出する。

$$AccuracyRate_{spam} = \frac{correct_{spam}}{test_{spam}} \quad (8)$$

$$AccuracyRate_{ham} = \frac{correct_{ham}}{test_{ham}} \quad (9)$$

6.3 併用の組み合わせによる結果の考察

併用 I の場合、一段階目の BONSAI で判定を行った際、正解率がスパムメール 96.9%, 正規メール 97.9% と共に高いので、正規メールと誤判定されるスパムメールが少ないが、少しの正規メールをスパムメールと誤判定してしまっている。二段階目の bsfilter は正規メールの正解率が高いので BONSAI がスパムメールと誤判定した正規メールを正しく判定している。しかし、スパムメールの正解率が低いので、BONSAI が正しく判定したスパムメールの

表 2 閾値変更前の bsfilter の判定結果

Table 2 Result of an accuracy rate of bsfilter before changing the evaluation value.

trainingmail	1000					
System	bsfilter[0.582]	BONSAI	併用 I	併用 II	併用 III	併用 IV
test _{total}	2000					
test _{spam}	1000					
test _{ham}	1000					
correct _{spam}	746	969	745	970	745	970
error _{spam}	254	31	255	30	255	30
correct _{ham}	1000	979	1000	979	1000	979
error _{ham}	0	21	0	21	0	21
AccuracyRate _{spam}	74.6	96.9	74.5	97.0	74.5	97.0
AccuracyRate _{ham}	100	97.9	100	97.9	100	97.9

うち、多くを誤判定してしまった。すなわち、二段階目で分類されるメールのほとんどはスパムメールであるため、スパムメールの判定は bsfilter に依存してしまい、低いスパムメールの正解率となってしまう。併用 II の場合、一段階目の BONSAI での分類では、BONSAI がスパムメールと判定したメールの中に少数の正規メールが含まれてしまった。一段階目の BONSAI がスパムメールと判定したメールに対しては、この時点で判定が決定するので、ここで誤判定された正規メールは回収することができない。二段階目の bsfilter で分類されるメールは、ほとんどが正規メールである。bsfilter は正規メールの正解率が高いので BONSAI が正しく判定した正規メールを回収することができた。また、BONSAI がほとんどのスパムメールを回収した後であるため、もともとスパムメールの正解率が低い bsfilter では正解率は大きく減少しなかった。併用 III の場合、一段階目の bsfilter での分類では、bsfilter は正規メールをすべて正しく判定したが、正規メールと判定したメールの中に大量のスパムメールが含まれてしまっている。一段階目で正規メールがすべて回収できたことから、二段階目の分類は混入するスパムメールを取り除くための判定を行うことになる。BONSAI がスパムメールを正規メールと誤判定してしまっただけで、わずかではあるが bsfilter のみに比べてスパムメールの正解率が減少した (74.6% から 74.5%)。併用 IV の場合、一段階目の bsfilter での分類では、bsfilter がスパムメールと判定したメールの中に正規メールは含まれなかった。しかし、bsfilter がスパムメールと判定したメール数は少ない。また、すべての正規メールが二段階目へ送られるため、BONSAI の分類性能ではすべての正規メールを回収することはできなかった。

以上 4 つの比較により、正規メールの正解率を可能な限り 100% に近づけるためには、正規

メールに対しての判定 2 回行う併用 I と III が有効であると考えられる。さらに、スパムメールの正解率を高くするためには併用 I と III の場合では、両方のフィルタが高いスパムメールの正解率が必要であるため、bsfilter のスパムメールの誤判定を減少させる必要がある。

6.4 メールスパム確率の分布

bsfilter のスパムメールの誤判定を減らすために、メール判定の閾値を従来より低く設定する。bsfilter のメール判定に使用される閾値は 0.582 以上でスパムメールに分類されるように設定されている³⁾。これは、正規メールの正解率を高くするため閾値を高く設定していると考えられ、その結果、多くのスパムメールを正規メールと判定している。つまり、閾値を下げることにより、スパムメールの正解率は上昇するが、反対に正規メールの正解率は減少してしまう。そこで、メールの閾値を変更するために各メールのスパム確率がどのように分布しているかを調べた。各メールのスパム確率の分布を表 3 に示す。変数 a を bsfilter の各メールに対してどの程度スパムメールに類似しているかを示す確率（メールのスパム確率）とする。スパムメール 1000 通、正規メール 1000 通の分類対象メール群について、スパム確率を 1 から 0 の間で 0.1 ずつ 10 個に区切り、各メールがどの範囲に含まれているかを示した。

表 3 メールスパム確率の分布
Table 3 Distribution of spam mail rate.

範囲	spam	ham
$0.9 \leq a \leq 1.0$	698	0
$0.8 \leq a < 0.9$	19	0
$0.7 \leq a < 0.8$	7	0
$0.6 \leq a < 0.7$	16	0
$0.5 \leq a < 0.6$	229	2
$0.4 \leq a < 0.5$	14	9
$0.3 \leq a < 0.4$	2	6
$0.2 \leq a < 0.3$	1	9
$0.1 \leq a < 0.2$	2	11
$0.0 \leq a < 0.1$	12	963

6.5 メール判定の閾値の変更

表 3 より、スパムメールの分布ではメールのスパム確率が 0.9 以上の範囲に含まれるメール数が 698 通と最も多い。正規メールの分布では、0.1 未満の範囲に含まれるメール数が 963 通と最も多く、正規メール全体の 96.3% を占めている。メール判定の閾値がデフォ

ルトの 0.582 では、多くのスパムメールが正規メールと判定されてしまい、スパムメールの誤判定を引き起こしている。そこで、スパムメールの誤判定を減少させるために閾値の値を従来よりも低く設定する。正規メールのほとんどがメールのスパム確率が 0.1 未満であり、スパムメールの誤判定をなるべく少なくする理由から、今回は閾値を 0.1 以上でスパムメールと判定することにした。最適な閾値の選択法については、今後の課題とする。

6.6 bsfilter の閾値変更後の結果

表 3 より、bsfilter の閾値を 0.1 以上でスパムメールと設定した場合、a が 0.1 未満である、スパムメール 12 通、正規メール 963 通が正規メールと判定され、a が 0.1 以上である、スパムメール 988 通、正規メール 37 通がスパムメールと判定される。正解率はスパムメールが 98.8%、正規メールが 96.3% となる。閾値変更前と比較して、スパムメールの正解率が 24.2% 上昇し、スパムメールの誤判定を大幅に減少させることができた。しかし、正規メールの正解率は閾値変更前と比較して 3.7% 減少していた。bsfilter のスパムメールの正解率を大きく上昇することができたため、併用 III を併用フィルターとして決定する。一段階目の bsfilter の分類では、スパムメールの正解率を高めることができたので、正規メールと判定されるスパムメールの数を減少させることができる。一方、正規メールの正解率は 100% ではないが、多くの正規メールを正しく判定できるので、BONSAI の正規メール、スパムメール共に正解率が高い特徴を利用して¹⁰⁾、bsfilter が誤判定した正規メールを回収できる可能性があり、かつ、スパムメールの正解率は大きく減少しないことが期待できる。

7. 併用フィルタリングの結果

閾値変更後の併用フィルタリングの分類結果を表 4 に示す。表 4 には、bsfilter の閾値を 0.1 以上でスパムメールと判定した結果、BONSAI を応用したメールフィルタリングで判定した結果、bsfilter と BONSAI を併用した結果を示す。

正規メールに関しては bsfilter の閾値を 0.582 以上でスパムメールと判定する基準から、0.1 以上でスパムメールと判定する基準に変更した結果、正解率は 96.3% となり、閾値変更前の 100% と比較して正解率が減少した。しかし、BONSAI と組み合わせることで、bsfilter が誤判定した正規メール 37 通中 36 通を正しく判定しており、99.9% という正規メールの正解率を得ていることから、BONSAI によって bsfilter の正規メールの誤判定を補う事ができた。bsfilter が誤判定したメールは、正規メールの中でメールのスパム確率が高いメールである。スパムメールの正解率は併用を行うことで bsfilter と BONSAI よりも低下したが、96.1% という正解率は、電子メール利用者にとって許容できる範囲ではないかと考えられる。

表 4 閾値変更後の bsfilter の判定結果

Table 4 Result of an accuracy of bsfilter after changing the evaluation value.

<i>trainingmail</i>	1000		
<i>System</i>	bsfilter[0.1]	BONSAI	併用フィルタ
<i>test_{total}</i>	2000		
<i>test_{spam}</i>	1000		
<i>test_{ham}</i>	1000		
<i>correct_{spam}</i>	988	969	961
<i>error_{spam}</i>	12	31	39
<i>correct_{ham}</i>	963	979	999
<i>error_{ham}</i>	37	21	1
<i>accuracy rate_{spam}</i>	98.8	96.9	96.1
<i>accuracy rate_{ham}</i>	96.3	97.9	99.9

8. 考 察

我々が提案するスパムメールフィルタは、併用するフィルタのそれぞれが異なる尺度で判定を行う事により、一方が誤判定してしまった正規メールをもう一方が正しく判定できる可能性を示している。正規メールおよび迷惑メールの両方の分類正解率の高い BONSAI を応用したメールフィルタの誤判定を減らし、利用者にとって最も重要な正規メールの回収率を上げるには、今回の実験から、次のような組み合わせと判定調節が有効だと考えられる。一段階目の bsfilter ではスパムメールの誤判定を可能な限り少なくすることが必要であるが、これを bsfilter の閾値を従来よりも低く設定することにより実現した。BONSAI を応用したメールフィルタを二段階目に配置することで、判定方式の違いから bsfilter が誤判定した正規メールを回収する事ができ、かつ、スパムメールの正解率が高いので、正規メールの高い正解率を維持しつつ、スパムメールの正解率を大幅に減少させないことが可能となった。電子メール利用者にとって最も重要なのは正規メールの正解率であるが、スパムメールの数が増加している現在では、スパムメールの正解率も重要となってきている。正規メールの判定に特に優れており、スパムメールの判定もある程度高いという点から併用フィルタリングは電子メール利用者のニーズに適したメールフィルタであると言える。また、今回は bsfilter との併用と行ったが、現在出回っているメールフィルタは一般的に正規メールの正解率を高くするように補正が行われており、BONSAI は単語の出現頻度と語順により判定を行う特殊なメールフィルタであるため、BONSAI と判定基準の異なる他の学習アルゴリズムと組み合わせることで、正規メールの分類正解率は BONSAI 単体比べて上昇すると考えられる。

9. おわりに

本稿では、BONSAI と bsfilter を組み合わせ、さらに bsfilter のメール判定に用いられる閾値を従来よりも低く設定することで正規メールとスパムメール共に、正解率の高いメールフィルタリング手法を検討した。併用フィルタの組み合わせ方の決定については、実際に 4 つの場合の結果を比較してそれぞれのフィルタの特徴について論じた。メール利用者にとって最も重要であるのは、正規メールを正しく判定する事であるため、正規メールを 2 回判定するようにした。閾値の変更についても、正規メールのほとんどのスパム確率が 0.1 未満であったことと、スパムメールの誤判定をなるべく少なく抑えるために、閾値が 0.1 以上でスパムメール判定されるように変更した。

今後は組み合わせの決定と bsfilter の閾値の最適化を行い、正規メールの正解率の維持とスパムメールの正解率の上昇を実現できるメールフィルタの開発を目指していきたい。

参 考 文 献

- 1) 安東 孝二, “世界の電子メールを spam 制御へ”. 情報処理, Vol.46, No.7, pp.741-746, 2005.
- 2) 景山 忠史, “spam メール の 現 状”. 情報処理, Vol.46, No.7, pp.747-751, 2005.
- 3) bsfilter-bayesian spam filter: <http://sourceforge.jp/projects/bsfilter/>
- 4) 杉井 学, 松野 浩嗣, “機械学習によるスパムメールの特徴の決定木表現”. 情報処理学会研究報告 2007-DPS-130(16), pp.183-188, 2007.
- 5) Shimozono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara, S., and Arikawa, S., “Knowledge Acquisition from Amino Acid Sequences by Machine Learning System BONSAI”, *Trans. Inform. Process. Soc. Japan*, Vol.35, pp.2009-2018, 1994.
- 6) Usuzaka, S., Kim, L., S., Tanaka, M., Matsuno, H., and Miyano, S., “A Machine Learning Approach to Reducing the Work of Experts in Article Selection from Database: A Case Study for Regulatory Relations of *S. cerevisiae* Genes in MEDLINE”, *Genome Informatics*, Vol. 9, pp. 91-101, 1998.
- 7) 小川 健司, 稲葉 宏幸, “記号と未知語の分布を用いたベイジアンスパムフィルタの提案”. 電子情報通信学会技術研究報告, 2009-IA-108(460), pp.209-212, 2009.
- 8) 田端利宏, “SPAM メールフィルタリング:ベイジアンフィルタの解説”, 情報の科学と技術, pp464-468, 56 巻 10 号, 2006.
- 9) TREC 2007 Public Corpus : <http://plg.uwaterloo.ca/~gvcormac/treccorpus07/>
- 10) 角 朝香, 日下野 隆謙, 杉井 学, 松野 浩嗣, “機械学習を応用したスパムメールフィルタリング手法の検討と評価”. マルチメディア通信と分散処理ワークショップ論文集, pp.201-204, 2009.