SiteWideBayesFeedback

How can my users feed back mail for the Bayesian learner?

If you want to set up site-wide use of Bayesian classification, you should set up a way for your users to send in misclassified mail to be "learned" from.

If you create mailboxes for false positives and false negatives, you can then run a cron job intermittently to learn all the mails in that mailbox as spam (or non-spam). Details on having your users redirect from mail clients to these mailboxes without mangling the headers are at ResendingMailWithHeaders.

Using Procmail with learning

For one approach, see ProcmailToForwardMail.

Submitting multiple messages at once

For those who want an "easier" way, and that also works with Outlook Express AND Outlook... (This also allows users to submit many mail pieces at once.) This maintains full headers and bodies. (Or as best as I can tell - someone tell me if I'm wrong!)

Create a *new* mail message in Outlook/Express. Resize the windows so that you can see both your new message as well as the main O/OE window. Select the messages you want to send as Spam or Ham (probably not both in the same message) and drag them "into" the new message. This will send all the messages as attachments to the main email.

As an admin, I like to review all these submissions to be sure they are really valid to submit to Bayes for training. I open the mail account with IMAP - usually using OE, and drag the appropriate attached messages to the IMAP folder(s) I want to feed to sa-learn. This allows you to review all messages before they're learned, and gives the users a pretty easy way to submit FP's & FN's as well as any other submissions you need.

(I also setup two "drop boxes" for mail - say qqqspam and qqqnospam - as re-iterated below, make them difficult to "guess" as you don't want spammers filling up your spam or ham drop.)

Using procmail to remove forwarding info

For MUAs (Like Netscape/Mozilla) that do a good job with keeping original headers intact, (almost) all you need to do is forward the email to the feedback account and strip off the header added by the forward. See BayesFeedbackViaForwarding for details.

Pine bouncing, for instance, adds headers like:

```
ReSent-Date: Wed, 27 Oct 2004 08:57:14 -0500 (CDT)
ReSent-From: My Name <myname@company.com>
ReSent-To: myname+spam@company.com
ReSent-Subject: !!! How can you refuse such a sexy plus size single community!
ReSent-Message-ID:
```

which may be stripped with an appropriate .procmailrc / sed stanza:

```
:0fw: splitmsg.lck
| sed -e '/^ReSent-/ d'
```

IMAP folders

Another option, and one that's easier for most users to use, is to set up two public IMAP folders on your IMAP server, one for MissedSpam, one for NotSpam.

Then ask your users to move messages that SpamAssassin misses into the MissedSpam folder, and copy messages that SpamAssassin marked incorrectly as spam into the NotSpam folder.

You can then run sa-learn from a cron job over those folders to update the Bayesian databases.

See RemoteImapFolder for more info.

How to set up site wide aliases on postfix where ham and spam can be sent for learning with Postfix

The cookbook is available at http://gtmp.org/doku.php/publications:sa-postfix-en, it works fine for postfix. Note: don't call your aliases spam and ham unless you want spammers to flood the ham box 🍔 - ideally, the aliases should not be directly accessible by inbound mail at all, there should be restrictions in place to only allow local users to send mail to those aliases.