Subject: Bayes Learning
Posted by Budd on Fri, 25 Aug 2006 18:15:57 GMT
View Forum Message <> Reply to Message

Ok, I need a definitive answer here... I have pretty much a basic installation of NST, and because we are mostly Outlook and Outlook Express users, not using an Exchange server, we have up to this point not done much to try teaching the bayes database.

However, though NST is killing a good amount of spam, we'd like it to get better, so I finally took a look and found some of the dbx converter programs out there, and can now collect some false negative spam messages, convert them to mbox, and perform the SA learning process.

My question, though, is if I'm supposed to only be teaching it messages that should have been spam, or are we supposed to teach it both spam and ham? Is it effective to only keep false negative spam messages and to run those through the sa-learn --spam process each week? And if we're supposed to also teach it ham, are we supposed to consider ham as only false positive messages, or do we just collect a random bunch of good mail messages and teach SA those as ham?

Thanks for your help, great product!

-Budd Wright
Pelco Solutions, LLC
www.pelcosolutions.com

Subject: Re: Bayes Learning
Posted by support on Mon, 28 Aug 2006 09:54:05 GMT
View Forum Message <> Reply to Message

> However, though NST is killing a good amount of spam, we'd like
> it to get better, so I finally took a look and found some of
> the dbx converter programs out there, and can now collect some
> false negative spam messages, convert them to mbox, and perform
> the SA learning process.
>
> My question, though, is if I'm supposed to only be teaching it
> messages that should have been spam, or are we supposed to
> teach it both spam and ham? Is it effective to only keep false
> negative spam messages and to run those through the sa-learn
> --spam process each week? And if we're supposed to also teach
> it ham, are we supposed to consider ham as only false positive
> messages, or do we just collect a random bunch of good mail
> messages and teach SA those as ham?
>
> Thanks for your help, great product!

For best results, you should teach any false positives (ham falsely classified as spam)

and any false negatives (spam falsely classified as ham) mails you get.
See also:
http://spamassassin.apache.org/full/3.1.x/dist/doc/sa-learn.html
http://wiki.apache.org/spamassassin/BayesInSpamAssassin
http://wiki.apache.org/spamassassin/BayesFaq

Subject: Re: Bayes Learning
Posted by Budd on Mon, 28 Aug 2006 18:15:44 GMT
View Forum Message <> Reply to Message

Thanks... however, I have a question on this. I've taught the database several hundred spam messages, definitely more than 200, and yet, when looking at the NST administrator's Maintenance tab, where the SA statistics are shown, the Words (tokens), spam tokens and ham tokens all still display "0". Why is this?

Thanks,

-Budd

Subject: Re: Bayes Learning
Posted by support on Wed, 30 Aug 2006 08:22:35 GMT
View Forum Message <> Reply to Message

Budd wrote:

> Thanks... however, I have a question on this. I've taught the
> database several hundred spam messages, definitely more than
> 200, and yet, when looking at the NST administrator's
> Maintenance tab, where the SA statistics are shown, the Words
> (tokens), spam tokens and ham tokens all still display "0".
> Why is this?

What command line exactly did you use to teach the Bayes database ?

Subject: Re: Bayes Learning
Posted by Budd on Thu, 31 Aug 2006 14:25:09 GMT
View Forum Message <> Reply to Message

The standard spam learn command line found in the batch file:

---

sa\sa-learn -C sa\ruleset --showdots --spam --mbox "C:\bayes*spam*.mbx"

I place our mbox files in that specified path, and everytime I run it, it runs successfully, telling me that it learned [x] number of tokens from [x] number of messages. (Not the exact message it gives me, but you get the point.)

Yet, the Maintenance tab's Words (Tokens), Ham Words and Spam Words fields still display 0 for each, even after clicking the refresh button and restarting the service...

Subject: Re: Bayes Learning
Posted by support on Thu, 31 Aug 2006 14:31:13 GMT
View Forum Message <> Reply to Message

Budd wrote:

> The standard spam learn command line found in the batch file:
>
> sa\sa-learn -C sa\ruleset --showdots --spam --mbox
> "C:\bayes*spam*.mbx"
>
> I place our mbox files in that specified path, and everytime I
> run it, it runs successfully, telling me that it learned [x]
> number of tokens from [x] number of messages. (Not the exact
> message it gives me, but you get the point.)
>
> Yet, the Maintenance tab's Words (Tokens), Ham Words and Spam
> Words fields still display 0 for each, even after clicking the
> refresh button and restarting the service...

Please send us your nospamtoday.cf and local.cf configuration files.