
Subject: Spam removal getting worse

Posted by [Mario Konst](#) on Mon, 07 Feb 2005 08:26:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

NST was working perfectly last year. After installing an update last year NST performance degraded. Maybe with the latest version I found where the problem lies.

On the maintenance tab it shows:

words : 166994

HAM : 12574

SPAM : 6153

After learning 153 SPAM mails it shows:

words: 167069

HAM : 12584

SPAM: 6153

It seems while learning SPAM, specific word are classified as HAM instead of SPAM.

Is this correct? I would imagine by learning SPAM , specific word are learned as SPAM tokens.

(batch)command for learning spam:

c:

cd "c:\program files\no spam today!\sa"

sa-learn -c ruleset --spam --mbox "C:/Documents and Settings/Administrator/Application Data/Thunderbird/Profiles/default/steoiizz.slt/Mail/post/spam

pause

Mario Konst

Subject: Re: Spam removal getting worse

Posted by [support](#) on Mon, 07 Feb 2005 11:07:41 GMT

[View Forum Message](#) <> [Reply to Message](#)

The most likely explanation for the performance changes last year were the changes in scoring introduced by version 3.0 of SpamAssassin.

SpamAssassin 3.0 was designed to decrease the likelihood of false positives.

I would not put it past sa-learn, that after learning spams, the number of ham tokens increases, for reasons hard to guess at.

Unless you have lots of wrong results from the Bayes tests (high scores for ham mail, leading to false positives, and low scores for spam, leading to false negatives), I would not worry.

Subject: Re: Spam removal getting worse

Posted by [Heidner](#) on Wed, 09 Feb 2005 07:09:26 GMT

The scoring points for the bayesian filter were lowered significantly. I think the points were dropped from nearly 5 for the 99% spam category down to about 1.9 points. For me that made a very big difference. (I had widened the threshold for spam detection so lowering the bayesian score opened the flood gates...)

I really did want the new ruleset from the newer Spamassassin, it includes support for SPF, etc... so my solution was to copy the scores for the bayesian filter into a local.cf file... (saved a copy also for backup) and used the prior bayesian filter scores. I also boosted the score for domains that are listed on the spamhaus lists. Between the two the catch without false positives climbed back up to nearly 80%+ (I have a large whitelist) If you excluded the whitelisted items - I think NoSPAM TODAY is hitting nearly 95% accurate blocking.

Subject: Re: Spam removal getting worse

Posted by [support](#) on Wed, 09 Feb 2005 11:24:13 GMT

> The scoring points for the bayesian filter were lowered
> significantly. I think the points were dropped from nearly 5
> for the 99% spam category down to about 1.9 points. For me
> that made a very big difference. (I had widened the
> threshold for spam detection so lowering the bayesian score
> opened the flood gates...)

The Bayes scores are back now to the SpamAssassin distribution scores. We had them increased, because we felt that the SpamAssassin developers were a little too cautious here. But after many users complained about false positives caused by the Bayes filter, we went back.

If you have a well-maintained Bayes database, it is probably still advisable to increase the scores. Add something like this to your local.cf file, using a text editor:

```
score BAYES_40 0.1
score BAYES_50 0.2
score BAYES_60 0.9
score BAYES_80 3.0
score BAYES_95 4.5
score BAYES_99 5.4
```
