

A Human Factors Approach to Spam Filtering

Robert Beverly

MIT CSAIL

rbeverly@csail.mit.edu

July 27, 2009

Conference on Email and Anti-Spam 2009

No spam classifier is perfect

Okay in other ML fields, e.g.

- Handwriting recognition, search engines, music recommendation, etc.

But with spam:

- Adaptable, adversarial inputs
- Complexion of dataset severely unbalanced
- High cost of false positives
- Getting from 99.9% to 99.999%

Fighting a losing battle?

No spam classifier is perfect

Okay in other ML fields, e.g.

- Handwriting recognition, search engines, music recommendation, etc.

But with spam:

- Adaptable, adversarial inputs
- Complexion of dataset severely unbalanced
- High cost of false positives
- Getting from 99.9% to 99.999%

Fighting a losing battle?

No spam classifier is perfect

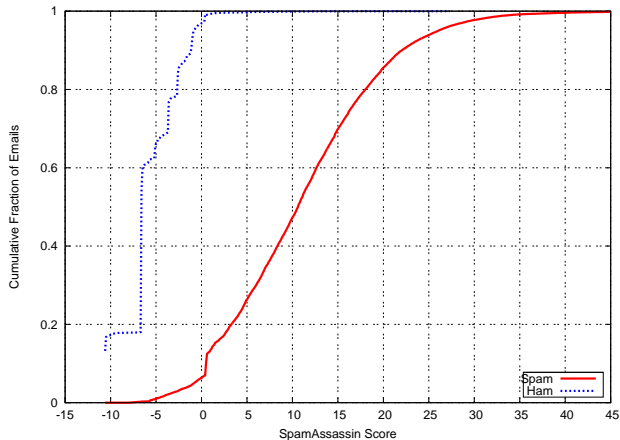
Okay in other ML fields, e.g.

- Handwriting recognition, search engines, music recommendation, etc.

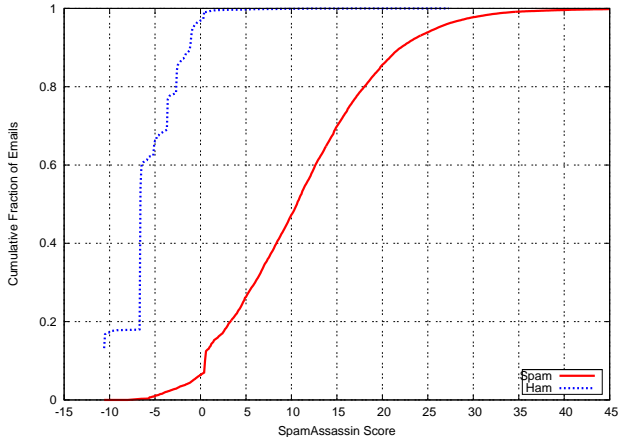
But with spam:

- Adaptable, adversarial inputs
- Complexion of dataset severely unbalanced
- High cost of false positives
- Getting from 99.9% to 99.999%

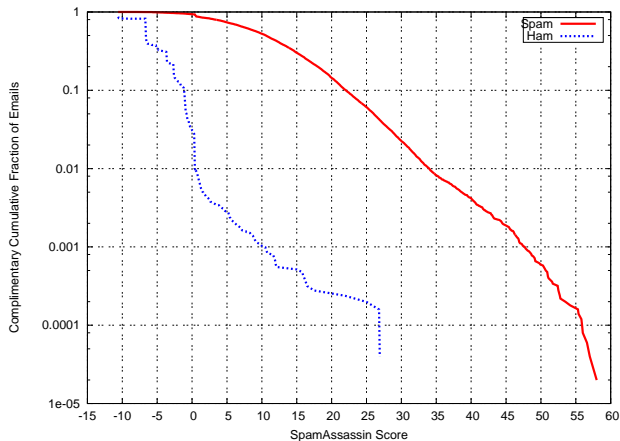
Fighting a losing battle?



- TREC 2007 dataset (~75k messages)
- Classified with SpamAssassin
- How close are mails to the threshold (5)?



- How close are mails to the threshold (5)?
- 99.72% of ham below threshold... good?



- No threshold gives zero FP/FN (well-known compromise)
- Deluge of spam implies this compromise is flawed
- 0.28% above \rightarrow 71 false positives

Approaching from a different direction...

The User Agent:

- Users interact with their email via a Mail User Agent (MUA), e.g. Outlook, Hotmail, etc.
- Note that besides going graphical, MUAs have changed little over past ~ 30 years
- Better incorporate human factors into a MUA

Human Factors Approach – Potential:

- 1 Make email more useful to the user
 - How are emails presented?
- 2 Humans ultimate arbiter of any mail's importance
 - How to better include, scale their decision process?
- 3 Remove burden of perfect classification from classifier
 - “good enough” filtering
- 4 Eliminate false positives

Innovate in the user agent

Human Factors Approach – Potential:

- 1 Make email more useful to the user
 - How are emails presented?
- 2 Humans ultimate arbiter of any mail's importance
 - How to better include, scale their decision process?
- 3 Remove burden of perfect classification from classifier
 - “good enough” filtering
- 4 Eliminate false positives

Innovate in the user agent

Position

- Separate *classification* from *filtering*

The inbox:

- Rethink the inbox: use a **single** mail folder, don't attempt to filter into spam, ham "folders"
- Use color, size, shade, order, and other *human factors* to present the inbox
- Presentation of email a function of *importance*

Proof-of-concept: SpamGUI Thunderbird extension...

Position

- Separate *classification* from *filtering*

The inbox:

- Rethink the inbox: use a **single** mail folder, don't attempt to filter into spam, ham "folders"
- Use color, size, shade, order, and other *human factors* to present the inbox
- Presentation of email a function of *importance*

Proof-of-concept: SpamGUI Thunderbird extension...

Spam & HCI

R. Beverly

The Problem

A Human
Factors
Approach

SpamGUI

Parting
Thoughts

Summary

Inbox for spanguit@berverly.net - Thunderbird

File Edit View Go Message Tools Help

Get Mail Write Address Book Reply Reply All Forward Tag Delete Junk Print Back

Subject	Sender	Spam
Re: chix	Rosalie Beverly, MD	-104.4
[Reuse] 2 free A/Cs, bedframe	nora10@zoragen.com	-5.7
Re: apartment switch is done	Beverly, Kelly	-1.9
Your Order with Amazon.com (#002-4659880-0...	Amazon.com Gift Cards	-1.7
Happy Easter from Parts4VWs.com	Parts4VWs.com	2
Your April SkyMiles STATEMENT	Delta Air Lines	2.1
support your darling sexuality	Robert Marshall	3.9
Send an Easter eCard to your family	FunCard	4.5
BeatriceTrahan sent you a message on Faceboo...	Facebook	5.2
MarciaEngland sent you a message on Facebook...	Facebook	6.8
Get your monthly free supply of Cialis	Viagra&Cialis Free	7.8
It not only will improve your health but also yo...	Marissa Sheffield	8.7
ascent your darling sexual times	Booker Davis	10.3
ED meds at lowest price - free!	Viagra&Cialis Free	11.2
Weekend	Tracy	12.7
itunes.com Invoice #54007	VIAGRA . Official Site	13.6
Cialis Super Active Tadalafil 20mg Only \$2.39 p...	Harvey W. Phipps	14.3
Весенняя охота на медведя	Камчатский МЕДВЕДЬ	14.3
Will you recognize old friend by photo)?	Conny Lemo	24.4
,,ä,Ä,Æ,µ,½%E,Ä,O,;E,ÄÄ"K,©,ä,æ	Kayla Guzman	27
fZfbfNfXfD,«É,É,É,cY"ñ,"S©,ß,Ä,Ä	Sakurai Kanae	37.7
Subject: [Reuse] 2 free A/Cs, bedfrar	From: nora10@zoragen	04/0

Hi all,

A Few Observations:

- A demarcation “line” naturally emerges to the eye, above which user (or UI) can ignore messages
- User part of filtering process, but only burdened by making spam decisions on a small number of emails around line
- Easy to scan for formerly false positive emails on the threshold border

Lots of work remains:

- No user studies performed yet
- Experimenting with several approaches

More generally:

- Users inundated with information, how can UI help?
- Spam is just one class of very unimportant information
- Lots of unused input “features;” systems designers should use them
- Learn best way to present email to user

Recognize that innovation is possible in the user agent

- We're fighting a losing battle trying to make spam classifiers perfect
- Separate act of classification from filtering
- As a community, think more about how HCI / human factors methods can help

Thanks!

<http://www.rbeverly.net/spamgui/>

Questions?