
Waarom Bayesiaans filteren de meest effectieve antispamtechniek is

Een detectiegraad van meer dan 98% dankzij een wiskundige methode

In dit white paper kunt u lezen hoe Bayesiaans filteren werkt en waarom dit de beste manier is om spam te bestrijden.

Inleiding

Dit white paper beschrijft hoe Bayesiaanse wiskunde kan worden gebruikt om spam tegen te gaan door middel van een adaptieve, 'statistisch intelligente', techniek met een zeer hoge spamdetectiegraad.

Tevens wordt uitgelegd waarom de Bayesiaanse benadering beter is dan relatief statische technieken zoals blacklistcontrole, vergelijkingen met databases van bekende spam en trefwoordcontrole. Hoewel deze technieken niet achterhaald zijn, zijn ze zonder Bayesiaanse filter niet betrouwbaar.

Inleiding	2
Bestaande spamdetectietechnieken	2
Hoe de Bayesiaanse spamfilter werkt.....	2
Waarom Bayesiaans filteren beter is.....	5
Over GFI MailEssentials.....	6
Over GFI.....	8

Bestaande spamdetectietechnieken

Spam is een steeds groter probleem. Het aantal spamberichten neemt met de dag toe – onderzoek heeft uitgewezen dat 50% van alle e-mail uit spam bestaat. Volgens de Radicati Group zal dit in het jaar 2007 zijn gestegen naar 70%. Bovendien is de werkwijze van spammers steeds geavanceerder en zijn spammers steeds beter in staat om de zogenaamde statische spambestrijdingsmethoden te verslaan.

De meeste antispamoplossingen maken gebruik van statische technieken. Dit betekent dat spammers de antispamsoftware gemakkelijk kunnen omzeilen door hun bericht een beetje aan te passen. Ze bestuderen de nieuwste antispamtechnieken en zoeken vervolgens naar manieren om deze te ontwijken.

Om spam op een effectieve manier te bestrijden is een nieuwe, adaptieve techniek nodig. Deze methode moet vertrouwd zijn met de door spammers gebruikte tactieken aangezien deze steeds veranderen. Ook moet de techniek zich kunnen aanpassen aan de organisatie die zij tegen spam beschermt. De oplossing zit in Bayesiaanse wiskunde.

Hoe de Bayesiaanse spamfilter werkt

Bayesiaanse filtering is gebaseerd op het principe dat de meeste gebeurtenissen met elkaar samenhangen en dat de kans dat iets in de toekomst zal gebeuren kan worden afgeleid uit eerdere gevallen. (Meer informatie over de wiskundige basis van Bayesiaanse filtering is te

vinden op Bayesian Parameter Estimation -

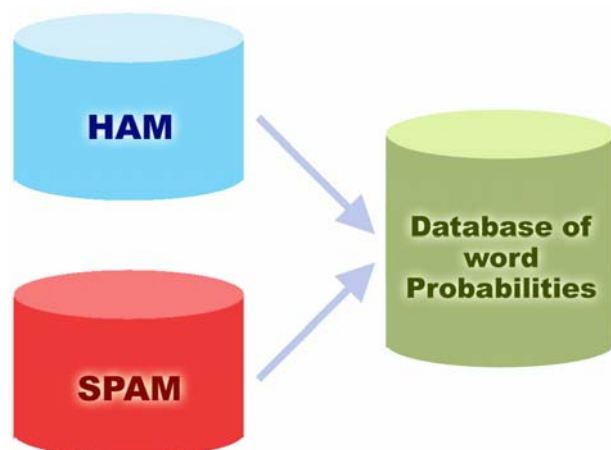
http://www-ccrma.stanford.edu/~jos/bayes/Bayesian_Parameter_Estimation.html

en An Introduction to Bayesian Networks and their Contemporary Applications - <http://www.niedermayer.ca/papers/bayesian/bayes.html>).

Dezelfde techniek kan worden gebruikt voor het classificeren van spam. Als een bepaald stuk tekst veel voorkomt in spam maar niet in legitieme mail, dan is het redelijk om aan te nemen dat dit bericht waarschijnlijk spam is.

Het creëren van een aangepaste Bayesiaanse begrippendatabase

Voordat met deze methode mail kan worden gefilterd, moet de gebruiker een database samenstellen met woorden en tekens (bijvoorbeeld het dollarteken, IP-adressen en domeinen, enzovoort) uit een sample van spam en legitieme mail (ook wel 'ham' genoemd).



Het creëren van een begrippendatabase voor de filter

Ieder woord of teken krijgt vervolgens een waarschijnlijkheidswaarde toegewezen. De waarschijnlijkheid wordt gebaseerd op berekeningen die rekening houden met hoe vaak een woord in spamberichten voorkomt en hoe vaak in legitieme mail (ham). Dit wordt gedaan door het analyseren van de uitgaande mail van de gebruikers en het analyseren van bekende spam. Alle woorden en tekens in beide groepen mail worden geanalyseerd om te berekenen hoe waarschijnlijk het is dat de aanwezigheid van een bepaald woord betekent dat het bericht een spambericht is.

Deze waarschijnlijkheid wordt als volgt berekend: als het woord "hypotheek" voorkomt in 400 van de 3000 spamberichten en in 5 van de 300 legitieme berichten, dan is de spamwaarschijnlijkheid 0.8889 (berekening: $[400/3000]$ gedeeld door $[5/300 + 400/3000]$).

Het creëren van de hamdatabase (aangepast aan uw bedrijf)

De analyse van ham mail wordt uitgevoerd op de mail van de organisatie, en is dus speciaal aangepast aan die organisatie. Een bank gebruikt bijvoorbeeld vaak het woord “hypotheek” en zou dus een groot aantal valse meldingen krijgen als er een algemeen hamdatabestand gebruikt zou worden. Als de Bayesiaanse filter zich eenmaal aan uw bedrijf heeft aangepast, houdt hij rekening met de uitgaande mail van het bedrijf (en herkent hij dus “hypotheek” als een woord dat vaak wordt gebruikt in legitieme berichten). Op deze manier wordt veel meer spam gedetecteerd en ontvangt u minder valse meldingen.

Antispamsoftware met hele beperkte Bayesiaanse capaciteiten (zoals de spamfilter van Outlook) of de Internet Message Filter in Exchange Server) creëert geen aangepast hambestand voor uw bedrijf, maar gebruikt een standaard hambestand. Hoewel een dergelijke filter niet eerst getraind hoeft te worden, zijn er twee belangrijke nadelen:

1. Het hamdatabestand is openbaar en kan dus worden gekraakt door professionele spammers en zo worden omzeild. Als uw databestand uniek is, heeft het geen zin om het te kraken. Er zijn bijvoorbeeld hacks beschikbaar waarmee de Microsoft Outlook 2003 spamfilter en de Exchange Server spamfilter gepasseerd kunnen worden.
2. Een algemeen hambestand is niet aangepast aan uw bedrijf en zal dus leiden tot een hoger aantal valse meldingen.

Het creëren van een spamdatabase

Naast ham maakt de Bayesiaanse filter ook gebruik van een spambestand. Dit spambestand moet een omvangrijk sample van bekende spam bevatten en moet continu door de antispamsoftware worden geactualiseerd met de nieuwste spam. Hierdoor bent u ervan verzekerd dat de Bayesiaanse filter op de hoogte is van de nieuwste trucs en dus een hoge detectiegraad bereikt (let op: het bijwerken vindt plaats na afloop van de leerperiode van twee weken).

Hoe de filtering plaatsvindt

Als de ham- en spambestanden eenmaal zijn gecreëerd, kunnen de waarschijnlijkheden worden berekend en is de filter klaar voor gebruik.

Wanneer een nieuw bericht binnenkomt wordt het afgebroken in woorden. De meest relevante woorden (de woorden die het belangrijkste zijn voor het bepalen of het om een spambericht gaat) worden eruit gevist. Uit deze woorden leidt de Bayesiaanse filter af hoe waarschijnlijk het is dat het nieuwe bericht een spambericht is. Als de waarschijnlijkheid hoger is dan een drempelwaarde (bijvoorbeeld 0,9), dan wordt het bericht als spam beschouwd.

De Bayesiaanse manier van spambestrijding is zeer effectief – volgens een artikel van de BBC uit mei 2003 kan een Bayesiaanse filter een spamdetectiegraad van meer dan 99,7% bereiken

met een zeer laag aantal valse meldingen!

Waarom Bayesiaans filteren beter is

1. De Bayesiaanse methode kijkt naar het gehele bericht. De Bayesiaanse filter herkent niet alleen trefwoorden die op spam wijzen, maar ook woorden die op legitieme e-mail wijzen. Bijvoorbeeld: niet iedere e-mail waar het woord “gratis” of “cash” in voorkomt is spam. Het voordeel van de Bayesiaanse methode is dat wordt gekeken naar de interessantste woorden (bepaald door de afwijking ten opzichte van het gemiddelde) en vervolgens wordt berekend hoe waarschijnlijk het is dat het om spam gaat. De Bayesiaanse filter let op woorden als “cash” en “gratis” maar herkent ook de naam van uw zakenrelatie die het bericht heeft gestuurd en classificeert het bericht vervolgens als legitiem. Doordat alle aspecten van een bericht in overweging worden genomen is Bayesiaans filteren dus een veel intelligentere benadering dan bijvoorbeeld het controleren van trefwoorden, waarbij mail als spam wordt geclassificeerd op basis van een enkel woord.
2. Een Bayesiaanse filter past zich continu aan – doordat de filter leert van nieuwe spam en nieuwe legitieme uitgaande mail, ontwikkelt hij zich en past hij zich aan aan nieuwe spamtechnieken. Bijvoorbeeld: toen spammers “g-r-a-t-i-s” gingen gebruiken in plaats van “gratis”, kwamen hun berichten door de trefwoordcontrole heen totdat “g-r-a-t-i-s” werd toegevoegd aan het trefwoordenbestand. De Bayesiaanse filter heeft dergelijke tactieken echter onmiddellijk in de gaten. Sterker nog: als het woord “g-r-a-t-i-s” wordt gevonden, dan is dat een nog betere indicatie dat het om spam gaat. Het is namelijk onwaarschijnlijk dat dit woord in een hambericht voorkomt. Een ander voorbeeld is het gebruik van het de term “5ex” in plaats van “Sex”. In een hambericht zou je “5ex” waarschijnlijk niet tegenkomen. De kans is dus groot dat het een spambericht is.
3. De Bayesiaanse filter past zich aan aan de gebruiker. Hij leert over de e-mailgewoonten van het bedrijf en begrijpt dus bijvoorbeeld dat het woord “hypotheek” op spam kan duiden als het bedrijf dat de filter heeft geïnstalleerd een bedrijf is dat auto's verkoopt maar niet als het een financiële instantie is.
4. De Bayesiaanse methode is meertalig en internationaal. Aangezien de Bayesiaanse antispamfilter zich aanpast, kan hij voor iedere gewenste taal worden gebruikt. De meeste trefwoordenlijsten zijn alleen beschikbaar in het Engels en zijn dus nutteloos in niet-Engelstalige landen. De Bayesiaanse filter houdt tevens rekening met bepaalde afwijkingen van de standaardtaal en de verschillende manieren waarop bepaalde woorden in verschillende regio's worden gebruikt, zelfs als in die regio's dezelfde taal wordt gesproken. Hierdoor kan de filter nog meer spam tegenhouden.
5. In tegenstelling tot een trefwoordenfilter is een Bayesiaanse filter niet gemakkelijk voor de gek te houden. Een geavanceerde spammer die een Bayesiaanse filter voor de gek wil houden kan ofwel minder woorden gebruiken die gewoonlijk op spam wijzen (zoals gratis,

Viagra, etc.) ofwel meer woorden gebruiken die gewoonlijk op legitieme e-mail wijzen (zoals de naam van een bestaande contactpersoon). Het laatste is onmogelijk aangezien de spammer van iedere ontvanger het e-mailprofiel nodig zou hebben - en een spammer kan nooit voor iedere ontvanger dit soort informatie verzamelen. Het gebruik van neutrale woorden (bijvoorbeeld het woord "openbaar") werkt niet aangezien deze in de uiteindelijke analyse genegeerd worden. Het afbreken van woorden die veel in spamberichten worden gebruikt, zoals "h-y-p-o-t-h-e-e-k" in plaats van "hypotheek", betekent dat er een grotere kans bestaat dat het bericht spam is, aangezien in een legitieme e-mail het woord "hypotheek" nooit als "h-y-p-o-t-h-e-e-k" zou worden geschreven.

Bayesiaanse filters of geactualiseerde trefwoordenlijsten?

Sommige soorten antispamsoftware downloaden regelmatig nieuwe trefwoordenbestanden. Hoewel dit natuurlijk beter is dan het niet updaten van trefwoordenlijsten, is dit in feite een relatief onbetrouwbare methode die gemakkelijk omzeild kan worden. Het downloaden van updates maakt het een klein beetje moeilijker, maar het systeem is niet zo doeltreffend als een Bayesiaanse filter.

Even geduld alstublieft

Als een Bayesiaanse filter op de juiste manier is geïmplementeerd en zich heeft aangepast aan uw bedrijf, dan is er geen betere manier om spam te bestrijden. Zijn er ook nadelen? Er is één nadeel, maar dat hoeft geen al te groot probleem te zijn: voordat u de Bayesiaanse filter kunt gebruiken en beoordelen, moet u hem minstens twee weken de kans geven om zich aan te passen aan uw bedrijf. Als u dit niet doet, moet u zelf de ham- en spamdatabases creëren. Dit is een vrij complexe taak. Het is dus beter om de filter de tijd te geven om zich aan uw bedrijf aan te passen. De Bayesiaanse filter wordt steeds effectiever naarmate hij meer over het e-mailgedrag van uw organisatie leert. Geduld is dus echt een schone zaak!

Dit is dus belangrijk om te onthouden bij het evalueren van verschillende soorten antispamsoftware. Als het product over geavanceerde, aangepaste Bayesiaanse analyse beschikt, kan het pas na een paar weken beoordeeld worden. Het is mogelijk dat minder geavanceerde antispamsoftware aanvankelijk beter presteert. Na een paar weken zal de Bayesiaanse filter echter beter presteren dan conventionele antispamfilters.

Over GFI MailEssentials

GFI MailEssentials for Exchange/SMTP biedt spambestrijding op serverniveau zodat u niet op iedere desktop antispamsoftware hoeft te installeren en bij te werken. GFI MailEssentials laat zich snel installeren en beschikt over een hoge spamdetectiekans met behulp van Bayesiaanse analyse en andere methodes. U hoeft niets te configureren en dankzij de automatische whitelists zijn er zeer weinig onterechte meldingen. Het product past zich automatisch aan uw e-mailomgeving aan zodat de spamdetectie voortdurend aangepast en verbeterd wordt. U kunt tevens spam naar de junk mail folders van uw gebruikers dirigeren. GFI MailEssentials voegt

ook belangrijke e-mailfuncties toe aan uw mailserver: disclaimers, rapportage, mailarchivering en –monitoring, servergebaseerde autoreplies en POP3 downloading. U kunt meer informatie en een gratis trialversie downloaden op <http://www.gfi.nl/nl/mes/>.

Over GFI

GFI is een toonaangevende ontwikkelaar van software voor netwerkbeveiliging, inhoudsbeveiliging en messaging. Dankzij bekroonde technologie, een agressieve prijsstrategie en een sterke focus op MKB-bedrijven helpt GFI bedrijven over de hele wereld om maximale continuïteit en productiviteit te bewerkstelligen. GFI is opgericht in 1992 en heeft kantoren in Malta, Londen, Raleigh, Hong Kong, Adelaide en Hamburg die wereldwijd meer dan 200.000 installaties ondersteunen. GFI is een kanaalgericht bedrijf met meer dan 10.000 partners over de hele wereld. GFI is ook een Microsoft Gold Certified Partner. Meer informatie over GFI is te vinden op <http://www.gfi.nl>.

© 2007 GFI Software Ltd. Alle rechten voorbehouden. De informatie in dit document geeft het standpunt van GFI weer betreffende de besproken onderwerpen op de datum van publicatie. Aangezien GFI moet reageren op veranderende marktomstandigheden, moet dit document niet als een toezegging van GFI worden geïnterpreteerd. Na de publicatiedatum kan de correctheid van de informatie niet worden gegarandeerd. Dit white paper dient puur ter informatie. GFI GEEFT IN DIT DOCUMENT GEEN ENKELE GARANTIE, EXPLICIET NOCH IMPLICIET. GFI, GFI EndPointSecurity, GFI EventsManager, GFI FAXmaker, GFI MailEssentials, GFI MailSecurity, GFI MailArchiver, GFI LANguard, GFI Network Server Monitor, GFI WebMonitor en de bijbehorende logo's zijn ofwel geregistreerde handelsmerken of handelsmerken van GFI Software Ltd. in de Verenigde Staten en/of andere landen. Alle product- en bedrijfsnamen in dit persbericht zijn mogelijk handelsmerken van hun respectievelijke eigenaren.

