

Por qué el filtrado Bayesiano es la tecnología anti-spam más efectiva

Consiguiendo más de un 98% de detección utilizando una aproximación matemática

Este documento describe cómo funciona el filtrado Bayesiano y explica por qué es la mejor forma de combatir el spam.

Introducción

Este documento describe cómo las matemáticas Bayesianas pueden ser aplicadas al problema de spam, resultando en una técnica adaptativa de “inteligencia estadística”, que logra una gran tasa de detección de spam.

También explica el por qué la opción Bayesiana es la mejor forma de resolver el problema de spam de una vez por todas, ya que sobrepasa los obstáculos que tienen la mayoría de tecnologías estáticas, tales como lista negra (blacklist), análisis, comparación con bases de datos de spam conocido y análisis de palabras clave. Estas tecnologías no son obsoletas, pero no se puede confiar en ellas sin el filtro Bayesiano.

Introducción	2
Técnicas actuales de detección de spam	2
Cómo trabaja el filtro spam Bayesiano.....	2
Por qué es mejor el filtrado Bayesiano.....	5
Acerca de GFI MailEssentials	7
Acerca de GFI	8

Técnicas actuales de detección de spam

Spam es un problema que crece a cada instante. El número de correos spam incrementa diariamente – estudios muestran que más del 50% de todos los correos electrónicos actuales son spam; el Grupo Radicati predice que este alcanzará el 70% para el 2007. Además de esto, los spammers se están volviendo más sofisticados y están constantemente arreglándoselas para sobrepasar los métodos estáticos para combatir el spam.

Las técnicas actualmente usadas por la mayoría de software anti-spam son estáticas, lo que significa que es relativamente fácil evadirla al modificar un poco el mensaje. Para hacer esto, spammers, simplemente examinan las últimas técnicas anti-spam y encuentran formas de burlarlas.

Para combatir efectivamente el spam, es necesaria una nueva técnica adaptativa. Este método debe ser familiar con tácticas que usan los spammers a medida que pasa el tiempo. También debe ser capaz de adaptarse a la empresa específica a la que protege de spam. La respuesta se encuentra en las matemáticas Bayesianas.

Cómo trabaja el filtro spam Bayesiano

El filtrado Bayesiano se basa en el principio de que la mayoría de los sucesos están condicionados y que la probabilidad de que ocurra un suceso en el futuro puede ser deducido de las apariciones previas de ese suceso. (Más información sobre las bases matemáticas del

filtrado Bayesiano está disponible en –

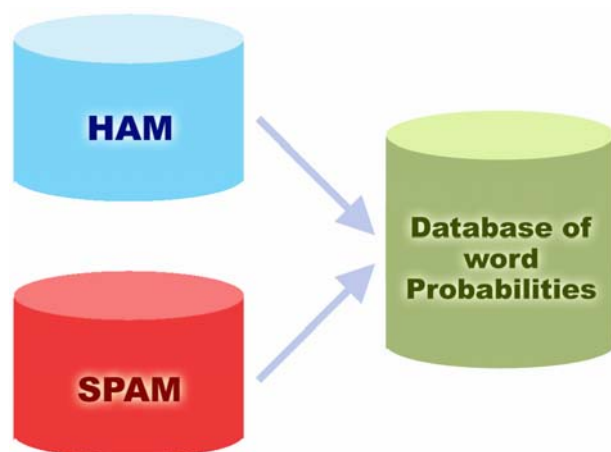
http://www-ccrma.stanford.edu/~jos/bayes/Bayesian_Parameter_Estimation.html

y una introducción a las redes Bayesianas y sus aplicaciones contemporáneas en - <http://www.niedermayer.ca/papers/bayesian/bayes.html>).

Esta misma técnica se puede utilizar para clasificar spam. Si algún patrón de texto se encuentra a menudo en el spam pero no en el correo legítimo, entonces sería razonable asumir que este correo es probablemente spam.

Creando una base de datos Bayesiana de palabras hecha a la medida

Antes de que el correo electrónico pueda ser filtrado utilizando este método, el usuario necesita generar una base de datos con palabras y testigos (cómo el signo \$, direcciones IP y dominios, etc), recogidos de un ejemplo de correo spam y de correo válido (referido como 'ham').



Creando una base de datos para el filtro

Se asigna entonces un valor de probabilidad para cada palabra o muestra; la probabilidad se basa en cálculos que tienen en cuenta que tan a menudo aparece la palabra en el spam frente al correo legítimo (ham). Esto se hace mediante el análisis del correo saliente de los usuarios y del correo spam conocido. Todas las palabras y muestras de ambos grupos son analizadas para generar la probabilidad de que una palabra concreta apunte que el correo sea spam.

Esta probabilidad de la palabra se calcula como sigue: Si la palabra "mortgage" aparece en 400 de 3.000 correos spam y en 5 de 300 correos legítimos, por ejemplo, entonces su probabilidad de ser spam sería 0,8889 (esto es, $[400/3000]$ dividido por $[5/300 + 400/3000]$).

Crear la base de datos de ham (a la medida de su empresa)

Es importante observar que este análisis del correo legítimo se realiza sobre el correo de la

empresa, y es por lo tanto hecho a la medida para esa empresa. Por ejemplo, una institución financiera podría utilizar la palabra "mortgage" más veces y obtendría muchos falsos positivos si utiliza un juego de reglas anti-spam general. Por otro lado, el filtro Bayesiano, si está hecho a la medida de su empresa mediante un periodo inicial de aprendizaje, toma nota del correo saliente válido de la empresa (y reconoce "mortgage" como frecuentemente utilizada en mensajes legítimos), y por lo tanto tiene mucho mejor ratio de detección de spam y mucho menor ratio de falsos positivos.

Observe que algunas aplicaciones anti-spam con capacidades Bayesianas muy básicas, como el filtro spam de Outlook o el Filtro de Mensajes de Internet de Exchange Server, no crean un archivo de datos del ham hecha a medida para su empresa, sino que incluye un archivo de datos ham estándar con la instalación. A pesar de que este método no requiere de un período de aprendizaje inicial, tiene dos defectos principales:

1. El archivo de datos ham está públicamente disponible y puede ser hackeado por spammers profesionales y por lo tanto evitado. Si el archivo de datos ham es único para su empresa, entonces el hacking del archivo de datos es inútil. Por ejemplo, hay hacks disponibles para evitar el filtro spam de Microsoft Outlook 2003 o de Exchange Server.
2. Este archivo de datos ham generalmente es uno, y por lo tanto como no está hecho a la medida de su empresa, no puede ser efectivo y usted sufrirá de un sensiblemente superior número de positivos falsos.

Crear la base de datos de spam

Además del correo ham, el filtro Bayesiano también se apoya en un archivo de datos spam. Este archivo de datos spam debe incluir un gran ejemplo de spam conocido y debe ser constantemente actualizado por el software anti-spam con lo último en spam. Esto asegurará que el filtro Bayesiano sea consciente de los últimos trucos spam, resultando en un alto ratio de detección (nota: este se adquiere una vez se finaliza el período de aprendizaje inicial de dos semanas).

Cómo está hecho el filtrado actual

Una vez han sido creadas las bases de datos de ham y spam, las probabilidades de las palabras pueden ser calculadas y el filtro está listo para su uso.

Cuando llega un nuevo correo, éste se descompone en palabras, y las más relevantes - es decir, aquellas que son más significativas para identificar si el correo es spam o no - son seleccionadas. De estas palabras, el filtro Bayesiano calcula la probabilidad de que el nuevo mensaje sea spam o no. Si la probabilidad es más grande que un umbral, digamos 0,9, entonces el mensaje se clasifica como spam.

Este acercamiento Bayesiano al spam es altamente efectivo – un artículo de la BBC de Mayo de 2003 informaba que los ratios de detección de spam de más de 99,7% pueden lograrse con

un muy bajo número de falsos positivos.

Por qué es mejor el filtrado Bayesiano

1. El método Bayesiano tiene en cuenta la totalidad del mensaje – Reconoce palabras clave que identifican el spam, pero también reconoce palabras que denotan correo válido. Por ejemplo: no todo el correo que contiene la palabra "free" y "cash" es spam. La ventaja del método Bayesiano es que considera la mayoría de palabras interesantes (definido por su desviación de la media) y da como resultado una probabilidad de que un mensaje sea spam. El método Bayesiano encontraría interesantes las palabras "cash" y "free" pero también reconocería el nombre del contacto de negocio que envió el mensaje y de ese modo clasificar el mensaje como legítimo, por ejemplo; esto permite que las palabras se "mantengan en equilibrio" entre sí. En otras palabras, el filtrado Bayesiano es una estrategia mucho más inteligente porque examina todos los aspectos de un mensaje, en oposición al análisis de palabras clave que clasifican un correo como spam en base a una sola palabra.
2. Un filtro Bayesiano están constantemente auto adaptándose – Mediante el aprendizaje del nuevo spam y la salida de nuevo correo válido, el filtro Bayesiano evoluciona y se adapta a las nuevas técnicas spam. Por ejemplo, cuando los spammers comenzaron a utilizar "f-r-e-e" en lugar de "free" consiguieron eludir los análisis de palabras hasta que "f-r-e-e" fue incluido en la base de datos de palabras. Por otro lado, el filtro Bayesiano advierte automáticamente estas tácticas; de hecho si se encuentra la palabra "f-r-e-e", incluso es un mejor indicador de spam. Otro ejemplo sería utilizar la palabra "5ex" en lugar de "Sex". Probablemente no tendrá una palabra 5ex en el correo ham, y por lo tanto la posibilidad de ser spam se incrementa.
3. La técnica Bayesiana es sensible al usuario – Aprende los hábitos de correo de la empresa y entiende que, por ejemplo, la palabra 'mortgage' podría indicar spam si la empresa que utiliza el filtro es, digamos, un distribuidor de automóviles, mientras que podría no indicar spam si la empresa es una institución financiera que trabaja con hipotecas.
4. El método Bayesiano es multilingüe e internacional – Un filtro anti-spam Bayesiano, al ser adaptable, puede utilizarse con cualquier idioma necesario. La mayoría de las listas de palabras clave sólo están disponibles en Inglés y son por lo tanto bastante inútiles en regiones no de habla Inglesa. El filtro Bayesiano también tiene en cuenta ciertas desviaciones del lenguaje o los diversos usos de ciertas palabras en áreas diferentes, incluso si se habla el mismo idioma. Esta inteligencia lo habilita como un filtro para atrapar más spam.
5. Un filtro bayesiano es difícil de burlar, a diferencia del filtro de palabras – Un spammer avanzado que quiera engañar a un filtro Bayesiano puede utilizar menos palabras 'malas' (es decir, palabras que habitualmente indican spam como free, Viagra, etc), o más

palabras que generalmente indican correo válido (como un nombre de contacto válido, etc). Haciendo lo último es imposible porque el spammer tendría que conocer el perfil de correo de cada destinatario - y un spammer nunca puede esperar obtener esta clase de información de cada destinatario deseado. Utilizando palabras neutras, por ejemplo la palabra "public", no funcionaría ya que estas son dejadas de lado en el análisis final. Utilizando palabras asociadas con el spam, como utilizar "m-o-r-t-g-a-g-e" en lugar de "mortgage", sólo incrementará la posibilidad de que el mensaje sea spam, ya que un usuario legítimo raramente utilizará la palabra "mortgage" como "m-o-r-t-g-a-g-e".

¿Filtros Bayesianos o listas actualizadas de palabras claves?

Algunos tipos de software anti-spam regularmente descargan nuevos archivos de palabras clave. Mientras que esto, por supuesto, es mejor que no actualizar las listas de palabras clave, el hecho es que es una aproximación un poco parcheada que puede ser fácilmente sobrepasada. El descargar actualizaciones, lo hace un poco más difícil, pero el sistema principal tiene imperfecciones comparado con el filtro Bayesiano.

¿Qué es lo que captura?

El filtrado Bayesiano, si está implementado correctamente y hecho a la medida de su empresa es de largo la tecnología más efectiva para combatir el spam. ¿Hay algún inconveniente? Bien, de alguna forma hay un inconveniente, pero puede ser fácilmente superado: Antes de poder utilizar y juzgar al filtro Bayesiano, tiene que esperar a que aprenda durante al menos dos semanas – eso o crear usted mismo las bases de datos de ham y spam. Esta tarea puede ser bastante compleja, por lo que es mejor esperar hasta que el filtro haya tenido tiempo de aprender. A lo largo del tiempo, el filtro Bayesiano se vuelve más y más eficaz ya que aprende más sobre los hábitos de correo de su organización. Para citar el antiguo dicho, buenas cosas le llegan al que espera.

Esto es importante de recordar al momento de evaluar un software anti-spam. Si el producto tiene avanzado, análisis Bayesiano personalizado, entonces puede ser juzgado después de unas pocas semanas. Es probable que software básico anti-spam pueda funcionar mejor inicialmente, pero después de unas pocas semanas, el filtro Bayesiano mejora y sobrepasa los filtros convencionales anti-spam de una vez por todas.

Acerca de GFI MailEssentials

GFI MailEssentials for Exchange/SMTP ofrece protección spam a nivel de servidor y elimina la necesidad de instalar y actualizar software anti-spam en cada escritorio. GFI MailEssentials ofrece una rápida puesta en marcha y una alta tasa de detección de spam utilizando análisis Bayesiano y otros métodos. No se requiere configuración, muy pocos falsos positivos mediante su lista Blanca automática, y la habilidad de adaptarse automáticamente a su entorno de correo electrónico, para así afinar constantemente y mejorar la detección de spam. También le permite organizar el spam en las carpetas de correo basura de los usuarios. GFI MailEssentials también agrega herramientas clave de correo electrónico a su servidor de correo: avisos corporativos, informes, archivo y monitoreo de correo electrónico, auto respuestas basadas en servidor y descarga POP3. Mayor información y una versión de evaluación completa están disponibles en <http://www.gfihispana.com/es/mes/>.

Acerca de GFI

GFI es un destacado desarrollador de software que proporciona una única fuente para que los administradores de red dirijan sus necesidades en seguridad de red, seguridad de contenido y mensajería. Con una galardonada tecnología, una agresiva estrategia de precios y un fuerte enfoque en las pequeñas y medianas empresas, GFI es capaz de satisfacer la necesidad de continuidad y productividad de los negocios que tienen las organizaciones en una escala global. Fundada en 1992, GFI tiene oficinas en Malta, Londres, Raleigh, Hong Kong, Adelaide y Hamburgo que soportan más de 200.000 instalaciones en todo el mundo. GFI es una empresa enfocada a canal con más de 10.000 partners en todo el mundo. GFI es también Microsoft Gold Certified Partner. Se puede encontrar más información sobre GFI en <http://www.gfihispana.com>.

© 2007 GFI Software. Todos los derechos reservados. La información contenida en este documento representa la visión del momento de GFI sobre lo discutido a la fecha de la publicación. Como GFI debe responder a las condiciones de los cambios del mercado, no debe ser interpretado como obligación por parte de GFI, y GFI no puede garantizar la exactitud de la información presentada después de la fecha de publicación. Este Documento Blanco solo tiene propósito informativo. GFI NO DA GARANTIA, EXPRESA O IMPLICITAMENTE, EN ESTE DOCUMENTO. GFI, GFI EndPointSecurity, GFI EventsManager, GFI FAXmaker, GFI MailEssentials, GFI MailSecurity, GFI MailArchiver, GFI LANguard, GFI Network Server Monitor, GFI WebMonitor y sus logotipos son marcas registradas o marcas de GFI Software en los Estados Unidos y/o otros países. Todos los nombres de producto o empresas mencionados pueden ser marcas registradas de sus respectivos propietarios.

