Inteligencia Artificial 17(58), 45-48
doi:10.4114/ia.v19i58.1154

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Resumen de tesis: Modelo para la optimización de la ejecución de filtros anti-spam.

D. Ruano-Ordás
Departamento de Informática, Escuela Superior de Ingeniería Informática, Universidad de Vigo. Campus Universitario As Lagoas S/N, 32004, Ourense.
drordas@uvigo.es

Resumen En los últimos años, se ha extendido y popularizado la combinación de distintas técnicas de filtrado para el desarrollo de sistemas anti-spam eficaces. No obstante, aunque la precisión lograda por estos modelos ha aumentado considerablemente, su utilización ha conllevado la aparición de nuevos retos como la necesidad de reducir el excesivo uso de los recursos computacionales, incrementar la velocidad de filtrado y ajustar los pesos empleados para la combinación de diversas técnicas de filtrado. Con el fin de lograr este objetivo se han refinado varios aspectos incluyendo: *(i)* el diseño y desarrollo de pequeñas mejoras técnicas que permiten incrementar el rendimiento global del filtro, *(ii)* la aplicación de algoritmos genéticos para aumentar la precisión de filtrado y *(iii)* el uso de algoritmos de planificación para mejorar la velocidad del sistema de clasificación.

Keywords: spam, rule-based systems, artificial intelligence, genetic algorithms, optimization, rule scheduling, e-mail.
Palabras clave: spam, sistemas basados en reglas, inteligencia artificial, algoritmos genéticos, optimización, planificación de reglas, e-mail.

1 Introducción

El establecimiento de la primera interconexión entre dos equipos remotos en 1969 originó el inicio de uno de los fenómenos tecnológicos más importantes de la humanidad, Internet. De hecho, se ha convertido en una parte esencial de la vida de muchas personas que residen en la mayoría de las naciones industrializadas alcanzando en 2016 un porcentaje de penetración del 46% de la población mundial [1].

Una de las razones que propiciaron la masiva proliferación de Internet se atribuye al servicio de correo electrónico (también conocido como e-mail). La comunicación fácil y rápida (casi instantánea) entre usuarios a través de mensajes de texto ha propiciado que este servicio adquiriese una sorprendente popularidad. Sin embargo, este hecho junto con la naturaleza incontrolable de Internet ha transformado las comunicaciones por e-mail en un marco para la promoción de anuncios ilegales (como los que se refieren a la venta de drogas), el envío de correos phishing, la propagación de virus y otras formas de fraude electrónico (también conocido como spam).

Gracias a los últimos avances alcanzados en los sistemas y tecnologías de la comunicación, los usuarios pueden disfrutar de acceso completo a Internet (24 horas al día) a través de dispositivos portables (ej. teléfonos inteligentes o tablets), fomentando el uso de sistemas de comunicación en tiempo real como los servicios de mensajería instantánea (IM).

A primera vista se podría considerar que las aplicaciones IM podrían reemplazar el uso de los servicios de e- mail tradicional debido a sus capacidades de comunicación en tiempo real y al intercambio de confirmaciones de entrega y lectura. Sin embargo *(i)* la capacidad de almacenamiento *(ii)* la accesibilidad universal y *(iii)* la posibilidad de adjuntar información adicional, transformaron el servicio de e-mail en un sistema crítico de comunicación tanto para uso particular como profesional. Los 4.087 billones de cuentas de e-mail disponibles en 2015 certifican una gran popularidad tanto en los entornos particulares (3.017 billones de cuentas) como en los empresariales (1.070

ISSN: 1137-3601, 1988-3064(on-line)

millones) [2]. Aunque el porcentaje de cuentas particulares representa más del 73% del total, los correos electrónicos enviados desde ellas sólo se corresponden con el 45.63% de los 193.6 billones de e-mails enviados cada día [3]. El tráfico masivo originado por las cuentas vinculadas a las empresas está motivado por el abandono de los sistemas de comunicación tradicionales debido a la simplicidad y velocidad ofrecida por el servicio de correo electrónico.

Este hecho ha propiciado que el tráfico de correos spam haya crecido exponencialmente [4], forzando a incrementar la efectividad y eficiencia de los servicios de filtrado spam para satisfacer la demanda en curso. Para tratar de solventar esta problemática, los frameworks de filtrado disponibles actualmente (ej. SpamAssassin [5]) permiten la combinación de varias técnicas anti-spam (ej. expresiones regulares, Naïve Bayes [6] ó Sender Policy Framework [7]) usando sistemas de clasificadores múltiples (MCS) [8].

Aunque esta aproximación introdujo mejoras significativas en la precisión de la clasificación, la combinación de múltiples técnicas ha causado también un gran incremento en el tiempo necesario para clasificar cada e-mail (tiempo de clasificación). Esta circunstancia, unida a la propagación a gran escala de los envíos spam, dificulta el desafío de mantener las bandejas de correo libres de spam. Para afrontar esta situación, esta tesis doctoral saca partido del desarrollo y aplicación de técnicas de optimización (ej. algoritmos genéticos o mecanismos de planificación de reglas) para reducir el tiempo de filtrado y asegurar la viabilidad del negocio del filtrado spam.

2 Evolución de la investigación

La tesis doctoral surgió de un proyecto de investigación y desarrollo gallego en 2010 [9]. Dicho proyecto estaba motivado por la necesidad de disponer de un sistema capaz de detectar y clasificar automáticamente el spam en plataformas Web 2.0.

Tal como se comentó anteriormente, en los últimos años, el uso de los servicios de Internet ha crecido más allá de lo que en un principio se podía imaginar, motivado principalmente por dos factores: *(i)* los últimos avances en las comunicaciones que han permitido estar conectado las 24 horas al día desde un simple dispositivo de mano y *(ii)* la aparición del concepto Web 2.0 que ofrece servicios como Wikis o Redes Sociales. El uso generalizado de estos servicios ha propiciado que la mayoría de los sitios web permitan a los usuarios darse de alta usando la misma información del perfil y credenciales utilizadas en las redes sociales como Facebook, Twitter o Google Plus. Esta nueva forma de inscripción se ha convertido en un fenómeno muy común, motivada principalmente por dos factores críticos: *(i)* la simplicidad y la comodidad que proporciona la posibilidad de inscribirse en un sitio web sólo mediante la inserción de la información de acceso del perfil de usuario en la red social y *(ii)* la falta de conciencia de los usuarios sobre el riesgo que supone el intercambio masivo de su propia información privada, como la dirección de correo electrónico. Esta circunstancia permitió a los spammers para obtener fácilmente los correos electrónicos de los usuarios y por lo tanto reactivar el negocio del spam.

Con el fin de obtener la voz acerca de la enorme importancia de esta situación, un informe de investigación reciente demuestra que el spam social se ha incrementado un 355% en el primer semestre de 2013 [6]. Por otra parte, los numerosos tipos de spam social existente a través de Internet (ej. link spam, text spam, o image-based spam) han impulsado la necesidad de actualizar las plataformas actuales de filtrado de correo no deseado con el fin de alcanzar tanto *(i)* la mejora de las técnicas anti-spam existentes y *(ii)* el desarrollo de nuevos métodos de filtrado de correos spam. Sin embargo, es importante tener en cuenta que el aumento de la complejidad de las plataformas anti-spam implica una reducción drástica del rendimiento de filtrado (mayor complejidad del filtro implica menor rendimiento de filtrado).

3 Estructura del trabajo

El trabajo realizado se sustenta en torno a tres contribuciones principales, las cuales han sido publicadas en revistas internacionales de impacto indexadas en el *Journal Citation Reports* (JCR).

La primera contribución presenta Wirebrush4SPAM [10], una nueva solución de filtrado capaz de clasificar los correos electrónicos con una mayor eficiencia y velocidad que las herramientas anti-spam más utilizadas (como SpamAssassin). El funcionamiento general de Wirebrush4SPAM está basado en el motor de reglas de SpamAssassin debido a su: *(i)* facilidad de uso, *(ii)* alto grado de configuración y finalmente, *(iii)* el elevado nivel de precisión alcanzado gracias a su habilidad para combinar distintas técnicas de filtrado anti-spam. Sin embargo, y tras un análisis detallado de SpamAssassin se identificaron una serie de procesos ineficientes que causan un impacto negativo en la velocidad del filtro. Para solventar estas carencias y obtener una herramienta capaz de clasificar correos hasta diez veces más rápido que su homólogo SpamAssassin, Wirebrush4SPAM incluye seis características básicas: *(i)* aprendizaje post-clasificación (LAR), *(ii)* evaluación inteligente del filtro (SFE), *(iii)* ejecución multi-hilo de las reglas que componen el filtro, *(iv)* reglas definitivas (SCR), *(v)* la inclusión de un framework de

planificación que permite crear y desplegar algoritmos de ordenación de reglas, y finalmente *(vi)* el uso del lenguaje de programación C debido a su gran velocidad de ejecución con un bajo uso de recursos computacionales.

La segunda contribución [11] define un marco de pruebas para medir las capacidades de optimización de diferentes algoritmos genéticos aplicados a incrementar la precisión de los sistemas de filtrado de correos spam (en términos de Falsos Negativos y Falsos Positivos). Se optó por usar algoritmos genéticos debido a la excelente relación entre el tiempo de convergencia necesario y las capacidades de optimización ofrecidas en otros dominios de aplicación. Concretamente se evaluó el rendimiento de (i) el algoritmo genético de objetivo único (SOEA) GrindStone4SPAM [12] y (ii) los algoritmos genéticos multi-objetivo (MOEA) NSGAII [13] y SPEA2 [14].

Actualmente, la gran cantidad de envíos spam junto con el alto grado de complejidad de las nuevas técnicas anti-spam ha propiciado la necesidad de mejorar continuamente el rendimiento de los sistemas de filtrado. Esta casuística motiva el trabajo realizado en la contribución que cierra el trabajo doctoral [15]. El objetivo principal consiste en definir e implementar varias estrategias de planificación de la ejecución de las reglas con el fin de reducir el tiempo requerido para filtrar e-mails y usar eficientemente los recursos computacionales disponibles. Además, se incluye un análisis comparativo de la velocidad alcanzada por cada esquema de planificación para determinar la estrategia más adecuada a la hora de mejorar el rendimiento del sistema filtrado.

4 Conclusiones y trabajo futuro

La tesis doctoral se ha realizado con el fin de solucionar los principales retos del filtrado de correos spam *(i)* el incremento de la velocidad de filtrado y *(ii)* el aumento de la complejidad de las técnicas de filtrado. Ambas problemáticas han causado un impacto negativo en los sistemas de filtrado forzando a que las compañías de filtrado adapten sus infraestructuras y servicios (tanto en equipamiento hardware como en software) para hacer frente a las necesidades actuales. Por ello, resulta indispensable destinar recursos a la investigación de nuevas técnicas y estrategias que permitan un avance significativo y continuo en el área.

La incorporación de diversas técnicas como SFE o LAR ha permitido optimizar el mecanismo de evaluación de todas las reglas y por tanto mejorar la eficiencia del sistema de filtrado. Por otro lado, se ha priorizado la entrega del resultado de clasificación frente a otras actividades inherentes al proceso de filtrado. Además, el uso de reglas *SCR* permite definir condiciones simples y únicas de clasificación (siempre y cuando sea posible la identificación de las mismas), mejorando con ello la velocidad de clasificación. Finalmente, se han identificado e implementado distintas heurísticas de planificación que permiten definir un nuevo orden de ejecución de las reglas, garantizando con ello una ejecución más rápida del proceso global de clasificación.

En lo que respecta a la mejora de la eficacia en clasificación, se han empleado distintos modelos (MOEA y SOEA) para la optimización de las puntuaciones asignadas a cada regla. Así, ha sido posible concluir que la utilización de NSGA-II para la optimización de distintos parámetros, resulta de gran utilidad en la mejora de la precisión de los filtros anti-spam.

Respecto al trabajo futuro, y motivado por los resultados obtenidos por los MOEA aplicados a la minimización de errores de tipo FP y FN (optimización bi-objetivo), se ha considerado la utilización de estos algoritmos sobre un mayor número de variables objetivo como: *(i)* el tiempo necesario para filtrar un mensaje o *(ii)* el número de reglas a evaluar, permitiendo eliminar así las reglas irrelevantes o redundantes.

Motivado por la naturaleza cambiante del spam (problema del *concept-drift*) y el continuo desarrollo de nuevas técnicas de spamming con el objetivo de evitar (o dificultar) la detección de los mensajes no legítimos, es indispensable: *(i)* la mejora de las técnicas anti-spam existentes y *(ii)* el diseño de técnicas anti-spam novedosas que permitan hacer frente a las nuevas estrategias empleadas por los spammers. Para ello, y teniendo en cuenta los prometedores resultados obtenidos (en términos de velocidad y precisión) por las técnicas de aprendizaje profundo (deep learning) en *(i)* la clasificación de documentos de texto y *(ii)* el reconocimiento e identificación de imágenes, resulta evidente destacar su aplicabilidad directa al dominio del spam, donde un e-mail no es más que un tipo especial de documento de texto que puede contener imágenes.

Actualmente la bioquímica y bioinformática son dos disciplinas esenciales a la hora de desarrollar y mejorar fármacos que permitan curar (o paliar) las diversas enfermedades del ser humano mediante la combinación de múltiples componentes químicos hasta obtener el fármaco más eficaz con el menor nivel de toxicidad posible. Debido a que es un proceso pseudo-manual (se realiza principalmente por ensayo y error), requiere un uso elevado tanto de recursos como de tiempo. Por ello, consideramos que gracias a la flexibilidad de Wirebrush4SPAM y los prometedores resultados alcanzados por los algoritmos genéticos, se podría diseñar de un sistema que se acople perfectamente a este ámbito de investigación y permita, de una manera sencilla y automática, evaluar la toxicidad de un fármaco partiendo de la proporción de cada uno de sus componentes químicos

Agradecimientos

Me gustaría agradecer principalmente la ayuda y apoyo de mi director, J.R. Méndez, sin el cual este trabajo doctoral no hubiese sido posible. También quiero destacar la ayuda de F. Fdez-Riverola, líder del grupo de investigación *Sistemas Informáticos de Nueva Generación* (SING). Finalmente, agradecer a la Universidad de Vigo la beca pre-doctoral otorgada que me ha permitido realizar esta tesis doctoral.

Referencias

[1] International Telecommunication Union (ITU). Manual for measuring ICT Access and Use by Households and Individuals. ITU Publications. 2014. ISBN: 978-92-61-14893-5

[2] Radicati S. Hoang Q. Email Statistics Report 2011-2015. Disponible en http://www.radicati.com/wp/wp-content/uploads/2011/05/Email-Statistics-Report-2011-2015-Executive-Summary.pdf [accedido Agosto 2016].

[3] Radicati S. Email Statistics Report, 2014-2018: Executive Summary. Disponible en: http://www.radicati.com/wp/wp-content/uploads/2014/01/Email-Statistics-Report-2014-2018-Executive-Summary.pdf [accedido Agosto 2016].

[4] Cisco Systems. Spam Overview – SenderBase: Global Volume. Disponible en: http://www.senderbase.org/static/spam/#tab=1 [accedido Agosto 2016]

[5] Apache Group. The Powerful #1 Open-Source Spam Filter. SpamAssassin. Disponible en: http://www.spamassassin.apache.org [accedido Agosto 2016].

[6] Metsis, V., Androutsopoulos, I., Paliouras, G. Spam filtering with Naive Bayes - Which Naive Bayes?. In Proc of The Third Conference on Email and Anti-Spam (CEAS'06), pp 27-28, Mountain View, California, USA.

[7] Görling S. An overview of the Sender Policy Framework (SPF) as an anti-phishing mechanism. Internet Research, 17:169-179, 2007. doi: 10.1108/10662240710737022

[8] Wózniak, M., Graña, M., Corchado, E. A survey of multiple classifier systems as hybrid systems. Information Fusion, 16:3-17. 2014. doi: 10.1016/j.inffus.2013.04.006.

[9] Sistema Aberto de Filtrado Anti-Spam para Xestores de Contidos Dixitais. Ref: 10TIC017E. Plan Galego I+D Suma. Disponible en: http://gain.xunta.es/repo/docs/3ef0262099f1aaaeeeb9a835ca8753eb.pdf [accedido Agosto 2016].

[10] Pérez-Díaz, N. Ruano-Ordás, D. Fdez-Riverola, F. Méndez, JR. Wirebrush4SPAM: a novel framework for improving efficiency on spam filtering services, Software Practice and Experience, 43:1299-1318. 2013. doi: 10.1002/spe.2135

[11] Yevseyeva, I. Basto-Fernandes, V. Ruano-Ordás, D. Méndez JR. Optimising anti-spam filters with evolutionary algorithms. Expert Systems with Applications, 40:4010-4021. 2013. doi: 10.1016/j.eswa.2013.01.008

[12] Méndez JR. Reboiro-Jato, M. Díaz, F. Díaz, E. Fdez-Riverola, F. Grindstone4Spam: An optimization toolkit for boosting e-mail classification. Journal of Systems and Software, 85:2909-2920. doi: 10.1016/j.jss.2012.06.027

[13] Deb, K. Pratap, A. Agarwal S. Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6:182-197. 2002. doi: 10.1109/4235.996017.

[14] Zitzer, E. Laumanns M. Thiele, L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization. In Proceedings of the Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN'2001).

[15] Ruano-Ordás, D. Fdez-Glez. J. Fdez-Riverola, F. Méndez JR. Effective scheduling strategies for boosting performance on rule-based spam filtering frameworks. *Journal of Systems and Software*, 86:3151-3161. 2013. doi: 10.1016/j.jss.2013.07.036.