

Lectura de la tesis de Enaitz Ezpeleta

21/10/2016

Programa De Doctorado: PROGRAMA DE DOCTORADO EN INGENIERÍA

En el pibunal de la tesis:

-Presidente: Dr. D. Manel Medina Llinas ([Universitat Politecnica de Catalunya](#))

-Vocal: Dr. D. Igor Santos Grueiro ([Universidad de Deusto](#))

-Vocal: Dr. D. Per Magnus Almgren ([Chalmers University of Technology](#))

-Vocal: Dr. D. José Ramón Méndez Reboredo ([Universidad de Vigo](#))

-Secretario: Dr. D. Iñaki Garitano Garitano ([Mondragon Unibertsitatea](#))

Resumen de la tesis:

Las campañas de correo electrónico no deseado, siguen siendo una de las mayores amenazas que afectan a millones de usuarios al día. Aunque las técnicas de detección de spam son capaces de detectar un porcentaje muy alto de spam, el problema está lejos de ser solventado, sobre todo por la cantidad tan alta de tráfico spam, y las nuevas estrategias utilizados por los atacantes.

Además el auge del número de usuarios de las redes social en Internet, muchos de los cuales publican mucha información privada en sus perfiles, han proporcionado que estos sitios se conviertan en atractivos para los atacantes, principalmente por dos razones: posibilidad de explotar la información guardada en los perfiles, y por el contacto directo con los usuarios mediante los perfiles, los grupos, las páginas... Como consecuencia, cada vez más actividades ilegales son detectados en estas redes.

Siendo los principales objetivos de los mensajes spam vender algo, crear alarma social, crear campañas de sensibilización, etc. este tipo de comunicaciones suelen tener una forma de escritura que puede ser usado para su detección.

Los principales objetivos de esta tesis son: demostrar que es posible desarrollar spam personalizado usando información publicada en redes sociales que eluda los sistemas actuales de detección; y diseñar y validar métodos nuevos para la detección y filtrado de spam usando técnicas de procesamiento de lenguaje natural. Además, estos sistemas deberán ser efectivos con el spam que se propaga dentro de las redes sociales.

Para conseguir el primer objetivo de la tesis se ha diseñado y desarrollado un sistema que permite enviar campañas de spam personalizado. Extraemos información pública de una famosa red social, el cual luego analizamos y lo usamos para crear plantillas de correo. Después se realizan diferentes experimentos en los que se envía spam típico y spam personalizado con el objetivo de comparar la eficiencia de cada tipo de spam.

En la segunda parte de la tesis presentamos tres nuevos modelos para el filtrado del spam. Estos métodos tienen como objetivo detectar



intencionalidad comercial no evidente en los textos que luego ayuden a clasificarlos. Para ello, planteamos que la utilización de técnicas de análisis de sentimiento y reconocimiento de personalidad pueden ser opciones interesantes. Tomando en cuenta eso hemos desarrollado los tres métodos: el primero utiliza análisis de sentimiento, el segundo reconocimiento de personalidad y el último usa combinadamente ambas técnicas.

Una vez que los métodos son descritos, se validan de forma experimental usando tres tipos diferentes de spam: email spam, SMS spam, y spam de una popular red social.