

Los spammers no piensan: usando reconocimiento de personalidad para el filtrado de spam en mensajes cortos

Enaitz Ezpeleta

Departamento de Electronica e Informatica
Mondragon Goi Eskola Politeknikoa
Mondragon Unibertsitatea
Email: eezpeleta@mondragon.edu

Urko Zurutuza

Departamento de Electronica e Informatica
Mondragon Goi Eskola Politeknikoa
Mondragon Unibertsitatea
Email: uzurutuza@mondragon.edu

José María Gómez Hidalgo

Pragsis Technologies
Email: jmgomez@pragsis.com

Resumen—El spam o correo basura sigue siendo un grave problema de seguridad, siendo responsable de un volumen muy alto de todos los correos electrónicos y comunicaciones entre personas. Además, el uso omnipresente de teléfonos inteligentes y redes sociales para comunicaciones cortas entre personas también atrae cada vez más spam y campañas ilegítimas. Éstas, además de tratarse de actividades ilegales, suponen una amenaza contra la privacidad de los usuarios. La gran mayoría de técnicas de filtrado de spam se enfocan en la clasificación automática de texto para generar modelos, muchas veces bayesianos, de detección. En este trabajo, los autores han optado por tratar de mejorar los modelos basados en el análisis estadístico de las palabras del contenido, utilizando técnicas de procesamiento de lenguaje natural. Tomando como referencia un conjunto de textos de Servicios de Mensajería Corta (*Short Message Service* ó SMS en inglés) etiquetados como spam/ham, se generan modelos de clasificación comunes de la literatura, siendo la entrada a los algoritmos de clasificación la frecuencia de las palabras que incluye cada mensaje y su etiqueta. Después, se analiza cada mensaje desde el punto de vista de la personalidad, añadiendo su puntuación al vector de características. Comparamos los resultados de los mejores clasificadores y filtros a través de los diferentes conjuntos de datos (con y sin personalidad) con el fin de demostrar la influencia de la personalidad. Los experimentos muestran que la función de personalidad mejora el filtrado spam de mensajes SMS, alcanzando una precisión del 98,94 %, y también la cifra de falsos positivos.

Palabras clave—Spam, Procesamiento de Lenguaje Natural, Reconocimiento de Personalidad

I. INTRODUCCIÓN

De la misma manera que el uso de teléfonos inteligentes y redes sociales está creciendo, el tráfico de mensajes cortos está aumentando en todo el mundo. Por ejemplo, más de seis mil millones de personas utilizaban un dispositivo móvil con capacidad de envío de mensajes SMS en junio de 2015, lo que significa que este tipo de mensajes puede llegar a alcanzar a más de otros tantos millones de usuarios [1]. De la misma manera, la famosa aplicación de mensajería instantánea WhatsApp, alcanzó los mil millones de usuarios en febrero del año 2016¹.

Este volumen de usuarios potenciales atraen también a usuarios malintencionados. Sin ir más lejos, los cuerpos de seguridad del estado desmantelaron una banda de delincuentes que afirmaban haber ganado al menos cinco millones de Euros durante la última década a través de campañas maliciosas de SMS de tarifas premium². Las campañas de mensajes maliciosos utilizando el sistema de comunicación por SMS resultan especialmente efectivas, con una tasa de apertura del mensaje del 98 % (siendo la tasa de apertura del marketing basado en correo electrónico un 22 %)³. En países como China o India, el spam en este medio es un problema reconocido con volúmenes de tráfico de un 20-30 %, y emergente en toda la zona de Oriente Medio y Asia [2].

Durante los últimos años, son varias las herramientas y sistemas que se han sido diseñadas y desarrolladas por investigadores para hacer frente a este problema. La mayoría de los estudios se centran en la clasificación automática de texto, generando vectores de características a partir de las propias palabras del mensaje. En este trabajo, se ha partido desde la hipótesis de que el reconocimiento de la intencionalidad de un mensaje puede ayudar a mejorar la clasificación, debido a que por lo general el spam se genera con la intención de realizar una acción comercial, de querer vender un producto o un servicio. Como paso intermedio, el análisis de la personalidad de cada mensaje puede introducir mejoras en la clasificación de los mensajes, ya que los mensajes de este tipo tienden a mantener las mismas características desde un punto de vista de personalidad (o comportamiento humano).

El objetivo principal de este trabajo es analizar la influencia de la personalidad en el filtrado de spam en mensajes cortos, y demostrar que las técnicas de reconocimiento de la personalidad pueden ayudar en esta tarea. Teniendo en cuenta bases de datos disponibles públicamente, en este trabajo nos centramos en los mensajes SMS, que son estructuralmente similares a otros mensajes cortos instantáneos. Mediante el uso de este conjunto de datos previamente etiquetado con las clases

¹<https://blog.whatsapp.com/616/One-billion/>

²<http://goo.gl/Gy9YHV>

³<http://goo.gl/CaxweY>

(spam/ham), calculamos la personalidad de cada mensaje y agregamos la función de la personalidad al conjunto de datos original, creando un nuevo conjunto de datos. Se realizan una serie de experimentos con el fin de encontrar el mejor clasificador para filtrar mensajes de spam, y se comparan los resultados de los mejores clasificadores y filtros a través de los diferentes conjuntos de datos (con y sin personalidad), analizando los resultados obtenidos.

El resto del artículo está organizado de la siguiente manera. La sección II analiza trabajos previos realizados en el área de las técnicas de filtrado de spam en mensajes cortos, e introduce al lector en las técnicas de reconocimiento de personalidad. La sección III describe el proceso de clasificación de mensajes y de agregación de una puntuación de personalidad. En la Sección IV se comparan los resultados obtenidos a partir de modelos de clasificación clásicos con los modelos de clasificación con la personalidad del mensaje como característica añadida. Finalmente, discutimos los resultados en la Sección V.

II. TRABAJOS RELACIONADOS

Tal y como se ha mencionado anteriormente, diferentes estudios relacionados con este tema se han llevado a cabo en los últimos años. En esta sección se presenta una introducción a los trabajos más relevantes sobre SMS spam y técnicas de reconocimiento de personalidad.

II-A. Spam en Servicios de Mensajería Corta

En los últimos años los grupos de cibercriminales han identificado los sistemas de mensajería corta como un buen lugar en el que desarrollar actividades maliciosas, atraídos por el alto número de usuarios que hacen uso de estos servicios. En este trabajo nos centramos únicamente en Servicios de Mensajería Corta, más conocidos por sus siglas en inglés SMS (*Short Message Service*). Este tipo de mensajes son estructuralmente similares a otro tipo de servicios de mensajería corta como WhatsApp o incluso Twitter, servicios con volúmenes muy altos de usuarios. La razón de escoger este tipo de mensajes es la de haber tenido acceso a un conjunto de datos de mensajes previamente etiquetados, que además permite a los autores comparar los resultados con trabajos previos, y a su vez la reproducción de los resultados por parte de terceros.

Delany y col. presentan en [3] una revisión de trabajos entorno al filtrado de spam SMS. Además, se discute sobre corpus disponibles públicamente y su disponibilidad para la investigación en el área. Almeida y col. [4] comparan diferentes métodos de aprendizaje automático, y remarcaron que las Máquinas de Vectores Soporte eran las que mejores resultados daban. Obtuvieron tasas de precisión de 97,64 % mediante el uso de este algoritmo. Además, ponían a disposición de la comunidad una colección de mensajes SMS donde se incluyen mensajes de spam.

En otros estudios recientes, como el de Narayan y Saxena [5] y el de Nagwani y Sharaff [6] utilizan clasificadores a dos niveles para obtener mejores resultados en el filtrado de

spam. Nuestra propuesta se centra en mejorar los resultados obtenidos por clasificadores de un único nivel.

II-B. Reconocimiento de Personalidad

La personalidad es un constructo o categoría descriptiva psicológica que tiene por objetivo explicar una gran variedad de comportamientos humanos por medio de unas pocas características individuales, medibles y estables [7]. Tal y como describen Celli y Poesio en [8], existen dos modelos que formalizan o categorizan la personalidad; el modelo de personalidad Myers-Briggs[9], que define la personalidad mediante las siguientes cuatro dimensiones: Extroversión o Introversión, Pensamiento o Emocional, Calificador o Perceptivo, y Sensorial o Intuitivo. El otro modelo, acuñado como *Big Five* [10], divide las dimensiones de la personalidad en rasgos o factores denominados como factor O (*Openness* o apertura a nuevas experiencias), factor C (*Conscientiousness* o responsabilidad), factor E (*Extraversion* o extroversión), factor A (*Agreeableness* o amabilidad) y factor N (*Neuroticism* o inestabilidad emocional), los cinco forman el acrónimo mnemotécnico "OCEAN".

Mairesse y col. muestran en [11] que cada texto contiene mucha información sobre la personalidad de los autores. Esa personalidad inherente es atractivo en el área del procesamiento de lenguaje natural. En los últimos años se han realizado investigaciones relativas al reconocimiento automático de personalidad en blogs [13], textos offline [11], o en redes sociales[14]. Shen y col. prueban en [12] que la predicción de la personalidad es factible en correos electrónicos, pudiendo predecirla con precisiones razonables. Este trabajo muestra que se puede extraer la personalidad a nivel de mensajes (y no de usuarios que los escriben), de modo que se puede utilizar en diferentes aplicaciones, como la que se describe en este trabajo.

III. DISEÑO E IMPLEMENTACIÓN DEL SISTEMA

La Figura 1 muestra el proceso en el que se ha llevado a cabo el trabajo. Por un lado se realiza un proceso de filtrado de spam en mensajes SMS mediante la selección de diez clasificadores, combinados cada uno con diferentes parámetros de ajuste, lo que da treinta modelos de clasificación. Por el otro lado, se computa el valor de personalidad de cada mensaje, agregando este valor al conjunto de datos original. A este segundo conjunto se le aplican las mismas combinaciones de clasificadores que en el original. Extrayendo los mejores clasificadores de este experimento, se realiza un nuevo experimento donde estos clasificadores se combinan con diferentes filtros y procesamientos en busca de los mejores resultados. Finalmente se comparan los resultados de los diez clasificadores que mejores resultados aportan tanto en el conjunto de datos original como en el conjunto donde la personalidad ha sido agregada. Los clasificadores se validan mediante la técnica validación cruzada en diez partes, donde el modelo se entrena con nueve partes y se valida con la décima. De este modo se obtienen medidas de precisión y de falsos positivos. La

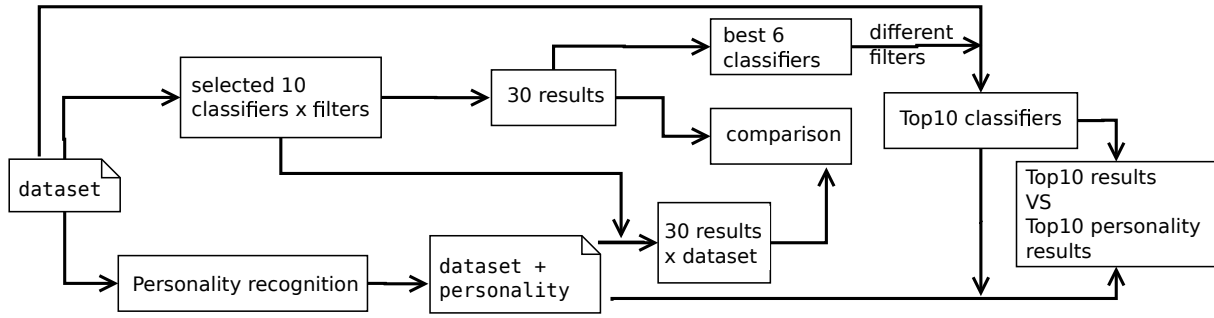


Figura 1. Proceso de análisis de personalidad en spam

precisión es el porcentaje de los ejemplos de la parte de testeo que han sido clasificados correctamente:

$$Precision = \frac{(VerdaderosPositivos+VerdaderosNegativos)}{Positivos+Negativos}$$

El conjunto de datos utilizado en el trabajo es *SMS Spam Collection v.1*⁴ (referido como SMSSpam a lo largo del trabajo). El conjunto de datos fue publicado en [4]. Está compuesto por 5.574 mensajes cortos reales y no codificados redactados en inglés, etiquetados como legítimos (ham) o spam. Concretamente, contiene 747 mensajes cortos spam y 4.827 mensajes cortos legítimos.

III-A. Reconocimiento de Personalidad

La primera fase del estudio aplica técnicas de reconocimiento de personalidad a cada mensaje SMS para después agregar este valor al conjunto de datos original. Para ello, se utiliza el modelo de personalidad Myers-Briggs. El modelo está compuesto por cuatro dimensiones (Extroversión o Introversión, Pensamiento o Emocional, Calificador o Perceptivo, y Sensorial o Intuitivo), necesarias para determinar el valor de personalidad de cada mensaje. Con el fin de computar estos valores, se ha utilizado un servicio Web conocido como *uClassify*⁵. Entre las diferentes opciones que dispone el servicio, se ha optado por las funciones Myers-Briggs desarrolladas por Mattias Östmar.

Tal y como indica el autor, cada función determina una dimensión del tipo de personalidad en concordancia con el modelo Myers-Briggs. El análisis se basa en el estilo de la escritura, por lo que no debe confundirse con los indicadores Myers-Briggs que determinan el tipo de personalidad basados en cuestionarios autocompletados. Los textos que utiliza la herramienta para ser entrenada están manualmente seleccionados, y están basados en los estilos de escritura y personalidad de Jensen y DiTiberio [15].

Estas son las siguientes funciones utilizadas:

- *Actitud Myers-Briggs*: Analiza la dimensión de Extroversión o Introversión .
- *Función de Juicio de Myers-Briggs*: Determina la dimensión de Pensamiento o de la Emocionalidad.

⁴<http://www.dt.fee.unicamp.br/tiago/smsspamcollection/>

⁵<https://www.uclassify.com>

- *Estilo de vida Myers-Briggs*: Determina el grado de Juicio o de Perceptividad.
- *Función de Percepción de Myers-Briggs*: Determina la dimensión Sensorial o la Intuitiva.

Cada función devuelve un número flotante del rango [0.0, 1.0] para cada par de características de la dimensión. Por ejemplo, si comprobamos un texto dado y obtenemos un valor X de Extroversión, el valor de la Introversión será de 1-X. De este modo, para cada mensaje utilizamos una única dimensión: Extroversión, Sentido, Juicio y Pensamiento. Estas cuatro variables se añaden al vector de características de cada mensaje SMS del conjunto de datos original.

III-B. Filtrado Spam

Para analizar la influencia de la personalidad en el filtrado de mensajes SMS, se seleccionan diez clasificadores representativos y filtros (o preprocesadores) que resultan adecuados para un problema de procesamiento de lenguaje natural. El criterio para seleccionar los clasificadores y los filtros se basa en un trabajo previo de los autores [18], y en los resultados de otro trabajo de investigación de Kumar y col.[19]. La siguiente lista muestra los algoritmos de clasificación utilizados en el trabajo:

- *Large-scale Bayesian logistic regression for text categorization*.
- *Discriminative parameter learning for Bayesian networks*.
- *Complement class Naive Bayes classifier*.
- *Multi-nominal Naive Bayes classifier*.
- *Updateable multi-nominal Naive Bayes classifier*.
- *Decision tree (C4.5)*.
- *Random tree*.
- *Forest of random trees*.
- *Support Vector Machine (SMO)*.
- *Adaptive boosting meta-algorithm with Naive Bayes*.

Además de los algoritmos arriba mencionados, se aplican los siguientes filtros a modo de preprocesamiento de texto:

- Un filtro que convierte una cadena de caracteres en un vector de palabras. Para ello, se combinan diferentes filtros: se pasan todas las letras a minúsculas; se eliminan caracteres especiales (.,:;\$& %=_@()?!+##[]) utilizando *tokenizers*; se generan n-gramas determinando máximos y mínimos para "n"; se aplica una técnica conocida como *stemming* para extraer la raíz de las palabras.

- Selección de atributos: se calcula la ganancia de información del atributo (palabra) respecto a la clase (spam/ham).

Mediante estas combinaciones, identificamos los diez mejores clasificadores parametrizados para la clasificación de spam SMS. Estos mismos son los utilizados con el conjunto que incluye el valor de la personalidad y comparar los resultados. La nomenclatura para los algoritmos y valores de los filtros utilizada se muestra en la tabla I.

	Significado
DMNB	DMNBtext
BLR	Bayesian Logistic Regression
CNB	Complement Naive Bayes
NBM	Naive Bayes Multinomial
NBMU	Naive Bayes Multinomial Updatable
RT	Random Tree
RF	Random Forest
SMO	Support Vector Machines
AB	Ada Boost with Naive Bayes
.c	idft F, tft F, outwc T
.i.c	idft T, tft F, outwc T
.i.t.c	idft T, tft T, outwc T
.stvw	String to Word Vector
.go	General options
.wtok	Word Tokenizer
.ngtok	NGram Tokenizer 1-3
.stemmer	Stemmer
.igain	Selección de atributos usando InfoGainAttributeEval

Tabla I
NOMENCLATURA UTILIZADA A LO LARGO DEL ARTÍCULO

IV. RESULTADOS EXPERIMENTALES

En esta sección se muestran los resultados obtenidos al aplicar el proceso de filtrado. Inicialmente, se lleva a cabo un análisis descriptivo de los datos, para después aplicar los algoritmos de clasificación y comparar resultados.

IV-A. Análisis Descriptivo

Una vez aplicado el servicio de reconocimiento de personalidad a cada mensaje, se considera interesante analizar los resultados obtenidos. La Tabla II muestra la cantidad y la proporción de mensajes tanto ham como spam catalogados en cada tipo de función de personalidad. Para este caso, *Ext.* se refiere a *Extroversión* y *Thi.* se refiere a *Thinking* o pensamiento.

	Total	Ext.	Sensing	Thi.	Judging
ham	4827	4392	3998	2431	1793
spam	747	599	566	238	431
Porcentaje(%)					
ham	100	91	83	50	37
spam	100	80	76	32	58

Tabla II
ANÁLISIS DESCRIPTIVO DE LA PERSONALIDAD DE LOS MENSAJES

Como se puede observar, hay diferencias significativas en todas las dimensiones del modelo de personalidad, pero sobre todo en las dos últimas dimensiones. Llama la atención observar que los mensajes de spam tienen menos personalidad

relativa al pensamiento, a la reflexión, y en cambio tienen más personalidad relativa al juicio o a la calificación. Estas diferencias hacen entrever que el valor de personalidad de un mensaje puede influir en los modelos de clasificación.

IV-B. Primer experimento: búsqueda de los mejores clasificadores de spam SMS

Tal y como se ha descrito anteriormente, se aplican los algoritmos seleccionados, combinados con los tres mejores filtros en base a [18]:

1. *stvw.go.wtok*
2. *i.t.c.stvw.go.ngtok.stemmer.igain*
3. *i.t.c.stvw.go.wtok*

La Tabla III muestra los resultados en términos de precisión (Prec.) y falsos positivos (FP), tanto para el conjunto de datos original, como para el conjunto de datos con la personalidad agregada como característica.

Nombre	Normal		Personalidad	
	FP	Prec	FP	Prec
SMO.3	3	98.73	5	98.64
NBM.3	12	98.69	4	98.71
NBMU.3	12	98.69	3	98.42
BLR.3	5	98.64	5	98.64
DMNB.1	10	98.62	5	98.73
BLR.2	2	98.60	2	98.60
NBM.1	23	98.53	14	98.62
SMO.2	4	98.53	3	98.56
NBMU.2	36	98.51	18	98.62
NBMU.1	29	98.49	19	98.64
CNB.1	31	98.44	18	98.60
NBM.2	52	98.37	44	98.51
DMNB.2	4	98.28	2	98.31
DMNB.3	4	98.28	2	98.31
CNB.2	64	98.19	54	98.35
CNB.3	56	98.17	23	98.73
BLR.1	1	97.45	0	96.23
SMO.1	0	97.45	0	97.43
J48.3	54	97.02	57	96.97
J48.2	58	96.90	60	96.86
J48.1	42	96.86	43	96.91
RF.2	0	96.38	0	96.47
RF.3	0	96.21	0	96.23
RT.1	25	95.60	22	95.39
RF.1	0	95.19	0	94.83
RT.3	84	95.16	86	94.94
RT.2	88	95.07	98	94.73
AB.2	167	91.44	167	91.44
AB.3	167	91.44	167	91.44
AB.1	188	91.32	188	91.32

Tabla III
COMPARACIÓN ENTRE RESULTADOS

Si analizamos la tabla de resultados (III), se puede observar que las máquinas de soporte vectorial dan los mejores resultados en términos de precisión, utilizando los parámetros del tercer filtro. En este caso, se observa que la personalidad no influye lo suficiente (ganancia de información) como para mejorar la precisión del modelo. Pero en otros dos casos, la personalidad utilizada en otros algoritmos ayuda a alcanzar el grado de precisión que el mejor algoritmo. Además, se puede ver que en la mayoría de los casos se mejora la

Name	Normal		Personalidad	
	FP	Acc	FP	Acc
NBMU.i.c.stwv.go.ngtok	28	98.85	26	98.83
NBMU.i.t.c.stwv.go.ngtok	27	98.82	19	98.94
NBM.i.t.c.stwv.go.ngtok	32	98.78	28	98.85
NBMU.i.t.c.stwv.go.ngtok.stemmer	23	98.78	19	98.80
NBM.c.stwv.go.wtok	13	98.76	5	98.78
NBM.i.t.c.stwv.go.ngtok.stemmer	34	98.76	30	98.78
NBMU.c.stwv.go.wtok	13	98.76	3	98.49
CNB.i.t.c.stwv.go.ngtok.stemmer	37	98.73	34	98.76
NBM.i.c.stwv.go.ngtok	37	98.73	33	98.78
NBM.i.c.stwv.go.ngtok.stemmer	36	98.73	33	98.76

Tabla IV
COMPARACIÓN DE LOS DIEZ MEJORES CLASIFICADORES

precision, o al menos se obtienen los mismos resultados. Otra observación interesante es aquella relacionada con la reducción generalizada de la tasa de falsos positivos cuando se aplica la personalidad. Es importante señalar que aunque tengan la misma precisión, reducir falsos positivos (aumentando falsos negativos) se considera como mejor resultado: es preferible que caiga spam en la bandeja de entrada (falso negativo) a que un mensaje legítimo sea considerado como spam.

Se demuestra por lo tanto que la personalidad puede ayudar a mejorar los resultados en el filtrado de spam SMS. No obstante, como último paso, se decide refinar la parametricación de los mejores algoritmos para buscar un nivel de precisión aún mayor.

IV-C. Mejora de resultados mediante combinaciones de filtros de procesamiento de texto

En este último paso se parte de los resultados del experimento anterior. En lugar de buscar los algoritmos que reporten los mejores resultados de clasificación, se trata de explorar las combinaciones de filtros de procesamiento de texto para los mejores algoritmos. De este modo, identificamos los seis algoritmos que mejores resultados dan de la Tabla III, y cada uno de ellos se combina con 56 ajustes de filtrado diferentes. Tras analizar los resultados sobre el conjunto de spam SMS original, se escogen los diez mejores en términos de precisión. Finalmente se utilizan los mismos clasificadores y ajustes de filtros con la característica de personalidad incluida, comparando los resultados en la Tabla IV.

La tabla muestra una mayor precisión con la combinación algoritmos y filtros que en el experimento anterior. Además, se evidencia que la característica de personalidad mejora la mayoría de resultados. En términos de precisión, se alcanza un 98,94 % de acierto en los test de validación cruzada, mejorando incluso el mejor de los resultados del conjunto de datos original. Solo en dos casos empeora la tasa de acierto al usar la personalidad, pero también en estos se mejora la tasa de falsos positivos. Esto último, es mejorado en todos los casos. Estos resultados validan los experimentos y las hipótesis planteadas al inicio del trabajo. Si tenemos en consideración que el conjunto de datos de partida es un conjunto relativamente reducido que contiene 875 mensajes cortos, esta pequeña variación podría suponer un volumen muy alto de mensajes

si generalizáramos los resultados a millones de mensajes (o miles de millones). Esto sí supone un importante avance en el campo.

V. CONCLUSIONES

Este trabajo plantea una hipótesis que nunca antes había sido explorada, y da pie a futuros trabajos para la mejora de los filtros de spam actuales. Los filtros de spam son actualmente herramientas comerciales, con un alto grado de madurez en el mercado, y basadas entre otras técnicas (como la reputación del emisor) en filtros o modelos de clasificación de texto. Los resultados dejan entrever que se pueden mejorar los filtros al añadir al conjunto de características variables relacionadas con el propio contenido o semántica del mensaje. Así, el análisis de sentimiento, o el reconocimiento de la personalidad de los mensajes suponen un avance hacia la detección de la intencionalidad.

En el artículo se muestra cómo añadiendo una característica de personalidad en cada mensaje de texto corto, puede mejorar el modelo de clasificador de spam tanto en términos de precisión y de falsos positivos. Utilizando el primer conjunto de datos se obtiene una tasa de acierto del 98.94 %. El número de falsos positivos se reduce en todos los casos de los diez mejores clasificadores. Aunque la diferencia en términos porcentuales es muy baja, un 0.01 porcentual sobre una base de un millón de mensajes implicaría acertar 10.000 mensajes más que el mejor de los clasificadores.

Finalmente, considerando que las funciones de reconocimiento de personalidad son independientes del texto (no se han utilizado mensajes SMS para entrenar un modelo de reconocimiento de personalidad), las conclusiones resultan muy positivas.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco bajo el proyecto de Investigación Básica y Aplicada SocialSPAM (PI_2014_1_102). Queremos agradecer a Mattias Östmar por las herramientas desarrolladas y publicadas, así como a Jon Kågström (fundador de uClassify⁶) por licenciar el uso de su API con motivos de esta investigación.

⁶<https://www.uclassify.com>

REFERENCIAS

- [1] Portio-Research-Limited, "SMS: the language of 6 billion people," June 2015.
- [2] Portio-Research-Limited, "Mobile Messaging Attacks Introduction," January 2011. *Trends and Examples*. Online <http://goo.gl/Q210eP>
- [3] S.J. Delany, M. Buckley, D. Greene, "SMS spam filtering: methods and data," *Expert Systems with Applications*, vol. 39, pp. 9899–9908, 2012.
- [4] T. Almeida, J.M. Gómez Hidalgo, A. Yamakami, "Contributions to the study of SMS spam filtering: new collection and results," en *Proceedings of the 11th ACM symposium on Document engineering*, pp.259–262, 2011.
- [5] A. Narayan, P. Saxena, "The curse of 140 characters: evaluating the efficacy of SMS spam detection on android," en *Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices*, pp.33–42, 2013.
- [6] N.K. Nagwani, A. Sharaff, "SMS spam filtering and thread identification using bi-level text classification and clustering techniques," *Journal of Information Science*, 2015.
- [7] A. Vinciarelli, G. Mohammadi, "A survey of personality computing," *IEEE Transactions on Affective Computing*, vol.5, n.3, pp.273–291, 2014.
- [8] F. Celli, M. Poesio, "Pr2: A language independent unsupervised tool for personality recognition from text," *arXiv preprint arXiv:1402.2796*, 2014.
- [9] I. Briggs Myers, P.B. Myers, "Gifts differing: Understanding personality type," *Consulting Psychologist's Press*, 1980.
- [10] A. Vinciarelli, G. Mohammadi, "Normal personality assessment in clinical practice: The NEO Personality Inventory," *Psychological assessment*, vol.4, n.1, pp.5, 1992.
- [11] F. Mairesse, M.A. Walker, M.R. Mehl, R.K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *J. Artif. Int. Res.*, vol.30, n.1, pp.457–500, 2007.
- [12] J. Shen, O. Brdiczka, J. Liu "Understanding email writers: Personality prediction from email messages," *User Modelling, Adaptation, and Personalization*, pp.318–330, 2013.
- [13] J. Oberlander, S. Nowson, "Whose Thumb is It Anyway?: Classifying Author Personality from Weblog Text," en *Proceedings of the COLING/ACL on Main Conference Poster Sessions (COLING-ACL'06)*, pp.627–634, Australia, 2006.
- [14] S. Bai and T. Zhu and L. Cheng "Big-Five Personality Prediction Based on User Behaviors at Social Network Sites," *CoRR*, abs/1204.4809, <http://arxiv.org/abs/1204.4809>, 2012.
- [15] G.H. Jensen, J.K. DiTiberio, "Personality and the Teaching of Composition," vol.20,1989.
- [16] P. Nakov, Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, T. Wilson "Semeval-2013 task 2: Sentiment analysis in twitter," 2013.
- [17] M.T. Nuruzzaman, C. Lee, D Choi "Independent and Personal SMS Spam Filtering," en *2011 IEEE 11th International Conference on Computer and Information Technology (CIT)*, pp.429–435, 2011.
- [18] E. Ezpeleta, U. Zurutuza, J.M. Gomez-Hidalgo "Does sentiment analysis help in bayesian spam filtering?," en *Hybrid Artificial Intelligent Systems: 11th International Conference, HAIS 2016, Sevilla, Spain, April 18-20, 2016*.
- [19] R.K. Kumar, G. Poonkuzhali, P. Sudhakar "Comparative study on email spam classifier using data mining techniques," en *Proceedings of the International MultiConference of Engineers and Computer Scientists*, pp.14–16. 2012.