Sánchez Fernández, Luis; Miñana Rontome, Tello Ismael; Arias Fisteus, Jesús; Basanta Val, Pablo; Fuentes Lorenzo, Damaris; Congosto Martínez, María Luz; Fernández García, Norberto, *Primeros resultados hacia la detección automática de bots en Twitter*. En: Payeras Capellà, María Magdalena; Ramis Bibiloni, Jaume (eds.) (2015). Jornadas de Ingeniería Telemática (JITEL 2015): libro de ponencias: del 14 al 16 de octubre de 2015, Palma de Mallorca. Universitat de les Illes Balears, 2015 (Pp. 191-196)
http://jitel15.uib.es/static/actas-jitel-2015.pdf

Primeros resultados hacia la detección automática de bots en Twitter

Luis Sánchez Fernández, Tello Miñana Rontomé,
Jesús Arias Fisteus, Pablo Basanta Val,
Damaris Fuentes-Lorenzo, Mariluz Congosto
Departamento de Ingeniería Telemática,
Universidad Carlos III de Madrid
Av. Universidad, 30, E-28911 Leganés (Madrid).
webtlab@inv.it.uc3m.es

Norberto Fernández García
Centro Universitario de la Defensa
Escuela Naval Militar
Plaza de España, s/n, 36920 Marín (Pontevedra)
norberto@cud.uvigo.es

***Resumen*—Las redes sociales en línea gozan de una gran popularidad. A su vez, es bien conocido que parte de la información que se difunde en dichas redes es, en muchos casos, de baja calidad, rumores o simplemente información falsa, difundida frecuentemente por cuentas automáticas o bots (aquellas que son gestionadas por programas de ordenador). En esta comunicación se presentan los primeros resultados de un trabajo que se está realizando para la detección de cuentas bot en Twitter. Entre los rasgos innovadores de este trabajo está el de utilizar indicadores sintácticos para analizar cuentas de Twitter.**

***Palabras Clave*—Redes sociales, Twitter, calidad de fuentes de información, bots**

I. INTRODUCCIÓN

Las redes sociales en línea [1] están suponiendo una revolución en la difusión y crecimiento de la información accesible a través de Internet adicional a la que supuso el propio desarrollo de la Web [2], [3]. Al igual que pasa con la Web, la naturaleza abierta de estas redes es una de las causas de su éxito. Sin embargo, también es una de sus debilidades: cualquiera puede difundir información en una red social y no hay ninguna garantía de que dicha información sea verídica [4], [5], [6]. En otras ocasiones, las redes sociales son utilizadas por organizaciones y empresas como medio para intentar influir en la opinión pública [7], [8], [9] o con fines publicitarios [10]. Desafortunadamente, también pueden darse situaciones en que las redes sociales sean una herramienta en la comisión de delitos [11].

En diferentes ámbitos (búsqueda de información [12], protección contra mensajes basura (spam) [13], seguridad [11], análisis de opinión [14], salud pública [15], [16], etc.), es interesante disponer de herramientas que puedan ayudar a evaluar la (buena o mala) calidad de fuentes de datos de redes sociales. Dentro de este marco, diferenciar las cuentas de usuarios humanos de aquellas que tienen detrás a programas de ordenador (llamadas bots) es interesante por varios motivos. Por una parte, en muchos casos los bots son productores de spam [17]. Por otra parte, cuando el objetivo es identificar usuarios humanos expertos en determinado tema es necesario desechar las cuentas bot.

Esta comunicación presenta los resultados de algunos experimentos que se han realizado con datos tomados de un conjunto de cuentas registradas en la red social Twitter, con el objetivo de detectar cuentas bot.

El resto del artículo está organizado de la siguiente forma. La sección II enuncia un conjunto de hipótesis cuya veracidad permitiría diferenciar automáticamente cuentas bot de cuentas de personas con una tasa de error baja. La sección III describe la metodología que se va a utilizar para validar dichas hipótesis. La sección IV describe los tipos de cuentas Twitter que se han analizado, así como las principales fuentes que se han utilizado para su selección. La sección V presenta los resultados que se han obtenido en los experimentos realizados. La sección VI presenta una somera discusión de dichos resultados. La comunicación termina presentando algunos trabajos relacionados y detallando las conclusiones obtenidas en este estudio, así como posibles líneas futuras de continuación del mismo.

II. HIPÓTESIS DE PARTIDA

La cuestión de distinguir el comportamiento de un ser humano del de una máquina es muy antigua. Un artículo clásico al respecto es el famoso "Computing machinery and Intelligence", de Alan Turin [18] (en realidad este trabajo plantea la cuestión inversa: si es posible construir una máquina cuyo comportamiento sea indistinguible del de los seres humanos).

Dado que nuestro primer objetivo, en el que nos centramos en esta comunicación, es el de ser capaces de distinguir automáticamente las cuentas Twitter de personas de las cuentas bot con una tasa de acierto elevada, el primer paso es construir de forma razonada algunas hipótesis cuya validez pudiera permitir diferenciar las cuentas bot de las que no lo son.

En primer lugar vamos a intentar identificar diferencias en el patrón temporal de tuiteo de las cuentas bot y de las cuentas de personas. Es bien conocido que el comportamiento temporal de los seres humanos presenta una importante variabilidad [19]. Es natural pensar que el comportamiento temporal de los programas de ordenador sea más regular (salvo que esté específicamente diseñado para no serlo). Esta idea ya ha sido explorada recientemente en trabajos como [20] y será la base de la primera de nuestras hipótesis.

En segundo lugar, nos planteamos explotar la dificultad de sintetizar lenguaje natural de forma automática [21]. Aparte de la complejidad intrínseca del lenguaje natural, es bien conocido que los seres humanos rompemos continuamente las reglas gramaticales del idioma en el que nos expresamos [21].

En resumen, al igual que en el caso anterior, esperamos encontrar una mayor variabilidad en el lenguaje expresado por los humanos que en el sintetizado automáticamente por un programa de ordenador.

Conviene en este caso matizar que existen experimentos previos que muestran que los programas de ordenador pueden de forma bastante efectiva producir lenguaje natural cuando mantienen un diálogo con un ser humano. En este contexto el programa de ordenador explota las sentencias generadas por el humano para producir las suyas propias. Uno de los primeros experimentos que se realizaron en esta línea fue el programa ELIZA [22]. De hecho, los bots conversacionales pueden ser considerados como una categoría de bots de Twitter en si misma.

La tercera hipótesis que vamos a plantear es específica del ámbito de las redes sociales en línea. Como ya se ha dicho, muchos de los bots se utilizan para difundir spam. En otros casos, se utilizan bots para difundir artículos de blogs y otras fuentes relacionados con diferentes temas. En todos estos casos es de esperar que los tuits de los bots contengan enlaces en un porcentaje elevado de casos, mientras que los humanos utilizaran enlaces con menos frecuencia.

En resumen de la discusión anterior, pretendemos evaluar las siguientes tres hipótesis:

H1 Las cuentas bot tuitean de forma regular, mientras que los humanos pueden estar largos intervalos de tiempo sin tuitear.

H2 Los tuits de las cuentas bot utilizan menos variedad en el lenguaje que los usuarios humanos.

H3 Los tuits de las cuentas bot contienen enlaces con una frecuencia mucho mayor que los tuits de las cuentas de personas.

En la sección siguiente se describe la metodología utilizada para intentar evaluar la validez de estas hipóteis.

III. METODOLOGÍA UTILIZADA

Para evaluar las tres hipótesis de partida presentadas en la sección anterior, se han definido los siguientes cuatro indicadores para una cuenta de Twitter:

1) Desviación típica de la diferencia entre segundos entre el sello temporal de un tuit y su anterior. La razón por la que se ha escogido este indicador es que si una cuenta tuitea de forma regular tendrá una desviación típica más baja que las cuentas en las que los intervalos de tuiteo sean irregulares. Relacionado con hipótesis **H1**.
2) Intervalo máximo entre un tuit y el anterior. Se espera que los bots tengan valores menores de este indicador porque tuitean de forma más regular. Relacionado con hipótesis **H1**.
3) El siguiente indicador está relacionado con la hipótesis **H2**. En relación con dicha hipótesis podrían contemplarse múltiples indicadores. En esta primera aproximación nos hemos conformado con uno solo. Partiremos de que las cuentas bot utilizan sentencias con una estructura sintáctica similar en sus tuits. Para intentar medir esto vamos a medir la variabilidad (en términos de desviación típica) de la profundidad del árbol sintáctico de cada sentencia escrita en una cuenta de Twitter. Para obtener dichos árboles sintácticos hemos utilizado el parser de la Universidad de Stanford [23].
4) Porcentaje de tuits que contienen enlaces. Derivado directamente de la hipótesis **H3**.

Para evaluar la validez de las hipótesis de partida, se han medido estos indicadores para un conjunto de cuentas en Twitter suficientemente grande como para que los resultados sean estadísticamente significativos y de las que se conoce a priori si son bots o no. A continuación los valores de los indicadores obtenidos para cada tipo de cuenta se han presentado en forma de box-plot, tal y como se describe en la sección de resultados.

Con el objetivo de que los datos recogidos en este trabajo puedan ser utilizados en otros experimentos posteriores, las cuentas no se han clasificado exclusivamente entre cuentas bot o no, tal y como se detalla en la sección siguiente.

IV. RECOLECCIÓN DE DATOS

Para realizar los experimentos se seleccionaron cuentas de Twitter de los siguientes tipos:

- Cuentas de medios de comunicación, todas ellas verificadas oficialmente por Twitter y obtenidas de [24].
- Cuentas de investigadores, profesores de universidad y periodistas, obtenidas de los seguidores de cuentas de universidades prestigiosas del mundo como Stanford y Cambridge.
- Cuentas de personajes famosos (políticos, actores, deportistas, etc.) y verificadas.
- Cuentas "fake": cuentas que parodian a los personajes de la categoría anterior, la mayoría de ellas obtenidas de [25], [26], [27].
- Cuentas bot, obtenidas de [28] y otras por medio de búsqueda manual.
- Cuentas aleatorias. La idea de esta categoría es disponer de cuentas de usuarios normales que no pertenezcan a ninguna de las categorías anteriores. Para obtener las cuentas aleatorias nos conectamos al `Stream API` de Twitter [29] y escogimos las primeras 40 cuentas en inglés y que hayan tuiteado al menos 10 veces desde que se creó la cuenta para tener algunos datos con los que poder realizar el análisis.

Para cada tipo de cuenta de Twitter se han analizado 40 cuentas. Esto debe ser suficiente, por ejemplo, para entrenar un clasificador con la técnica clásica del análisis discriminante de Fisher [30], [31] utilizando muestras de 20 dimensiones [32].

Como tenemos 6 tipos de cuentas y 40 cuentas de cada tipo, en total se han analizado 240 cuentas.

V. RESULTADOS OBTENIDOS

Cada uno de los cuatro indicadores descritos en la sección III se ha calculado para cada una de las 240 cuentas consideradas. Los resultados obtenidos se presentan por medio de un diagrama de cajas o box-plot [31], [33].

La figura 1[1] representa un diagrama de caja. Un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un

[1]"Boxplot" de Jumanbar - Trabajo propio. Disponible bajo la licencia CC BY-SA 3.0 vía Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Boxplot.svg#/media/File:Boxplot.svg

rectángulo, la "caja", y dos brazos, los "bigotes". La información que suministra el diagrama de caja, como se puede ver en la figura 1 es la siguiente:

- Primer cuartil (Q1)
- Segundo cuartil o mediana. Es la raya dentro de la caja.
- Tercer cuartil (Q3)
- Los bigotes, las líneas que se extienden desde la caja, se extienden hasta los valores máximo y mínimo de la serie o hasta 1.5 veces el Rango Inter Cuartílico (Q3 - Q1).
- Cuando los datos se extienden más allá de esto, significa que hay valores atípicos en la serie, que se marcan con pequeños círculos.

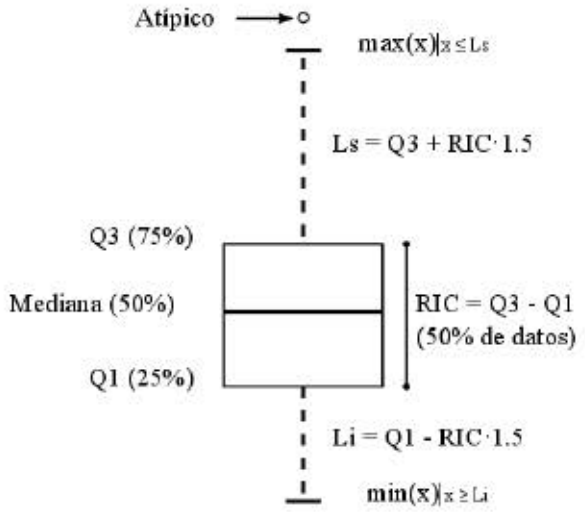

Fig. 1. Diagrama de caja o box-plot

Los resultados obtenidos para cada uno de los indicadores se presentan en las figuras 2, 3, 4 y 5. En cada una de estas figuras, el significado de las etiquetas en el eje de abscisas es el siguiente:

- X1: cuentas de medios de comunicación.
- X2: cuentas bot.
- X3: cuentas fake.
- X4: cuentas de famosos.
- X5: cuentas de investigadores, profesores y periodistas.
- X6: cuentas aleatorias.

Se debe tener en cuenta además que los resultados de las figuras 2 y 3 están expresados en escala logarítmica (logaritmo neperiano).

VI. DISCUSIÓN

Analizando los resultados obtenidos con los indicadores 1 y 2 se concluye que la hipótesis **H1** no es válida o al menos que dichos indicadores no miden adecuadamente la diferencia en el comportamiento temporal de los patrones de tuiteo de los usuarios humanos frente a los bots. Sí se observa en estos dos indicadores y especialmente en el indicador 1 una diferencia significativa entre el patrón de tuiteo de los medios de comunicación (estas cuentas pueden en muchos casos ser gestionadas total o parcialmente de manera automática) y el de las cuentas de usuarios humanos.

El indicador 3 en cambio confirma la validez de la hipótesis **H2** y de hecho muestra un comportamiento muy diferenciado

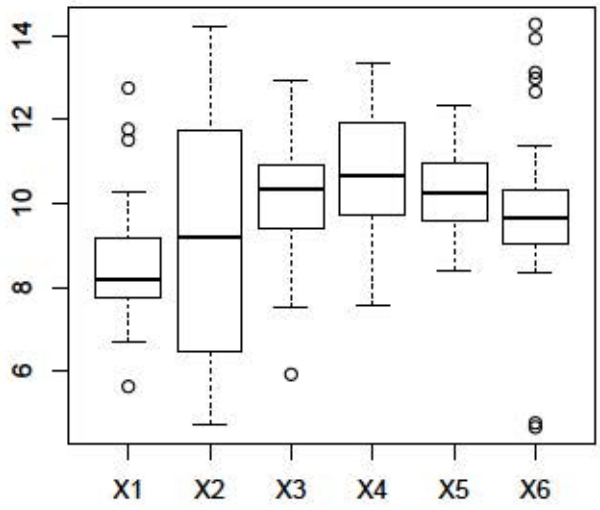

Fig. 2. Resultados para el indicador 1: desviación típica de la diferencia entre segundos entre dos tuits

entre las cuentas bot y las de los demás tipos. Llama la atención particularmente el buen comportamiento de este indicador para diferenciar cuentas bot de las de medios de comunicación, frente a lo ocurrido con los otros tres indicadores. Una posible explicación de este buen comportamiento podría ser la siguiente. Las cuentas Twitter de medios de comunicación habitualmente tuitean los titulares de las noticias que se van generando en el propio medio de comunicación. El texto de estos titulares es generado por periodistas y lógicamente presenta una variabilidad importante. Este indicador funciona mejor que los otros 3 considerados porque aunque las cuentas de medios de comunicación presentan similitudes con las cuentas bot en cuanto a patrón temporal y enlaces son muy diferentes desde el punto de vista de los aspectos lingüísticos de los tuits. En resumen, el indicador 3 se podría considerar adecuado para formar parte de un conjunto de indicadores para la detección de bots.

Finalmente, el indicador 4 también confirmaría la hipótesis **H3** (aunque de forma menos significativa que en el caso anterior). Excluyendo a las cuentas de medios de comunicación (que como ya se ha dicho muestran ciertas afinidades con las cuentas bot), se observa que más de la mitad de las cuentas bot tienen un porcentaje de tuits con enlaces próximo al 100% y notablemente más alto que la inmensa mayoría de las cuentas de los demás tipos. Por lo tanto, este indicador, si bien por sí solo no parece suficiente para distinguir cuentas bot de las que no lo son, parece adecuado para formar parte de un conjunto de indicadores que se utilizasen para detectar cuentas bot.

VII. TRABAJOS RELACIONADOS

Existen varios trabajos relacionados con el estudio de la fiabilidad de la información difundida en redes sociales.

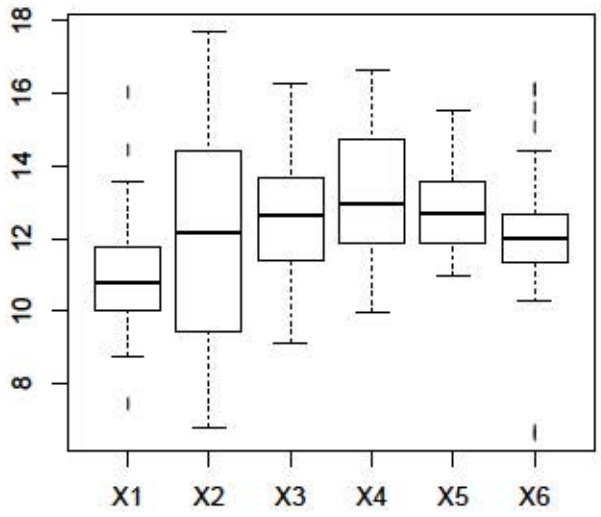

Fig. 3. Resultados para el indicador 2: Intervalo de tiempo máximo entre un tuit y el anterior

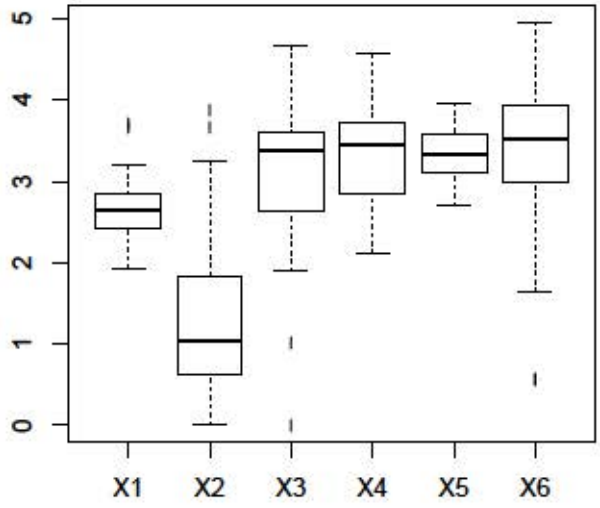

Fig. 4. Resultados para el indicador 3: Desviación típica de la profundidad del árbol sintáctico

En [34] se analiza la fiabilidad de los tuits difundidos durante el desastre de Fukushima. Este estudio se centra en los denominados tuits sintetizados, en los que un usuario sintetiza información proveniente de cierta fuente de información (en contraposición a un retuiteo puro) y que utilizaban el hashtag `#fukushima`. Se encontró que el 70% de los tuits provenian de fuentes de información fiables, con un porcentaje más bajo cuando los tuits provenían de cuentas con un perfil anónimo.

Otro trabajo en la misma línea es el de [4]. En este trabajo se analizó la difusión de tuits en el terremoto de Chile de 2010. En este estudio se concluyó que los rumores en Twitter tienden a ser más cuestionados que las noticias verídicas, lo que puede ser un instrumento en si mismo para su detección.

La detección de spammers puede considerarse como un apartado dentro del análisis de la fiabilidad de fuentes de información que ha sido estudiado en numerosos trabajos [13], [35], [36], [37].

En cuanto al caso concreto de la detección de bots en Twitter, existen algunos trabajos relacionados con lo aquí expuesto. En [38] se utilizan técnicas de aprendizaje máquina para distinguir spam bots de otros bots. El trabajo presentado en [20] se centra en identificar bots en base a patrones temporales. Principalmente, se han considerado dos aspectos de dichos patrones: 1) la distribución de los intervalos temporales entre 2 tuits; y 2) la frecuencia de tuiteo dependiendo de la hora del día (se debe tener en cuenta que este segundo aspecto solamente es útil cuando se dispone de información acerca de la zona horaria desde la que se tuitea). En [39] se propone una herramienta para la detección de bots. Entre los indicadores considerados en este trabajo se utilizan parámetros temporales. Se concluye que los bots tienden a tuitear en ráfagas, o con patrones temporales con una mayor regularidad que los humanos. Curiosamente en ese estudio se afirma que los bots no siempre tuitean con más frecuencia que los humanos. Entre otros indicadores también utilizados en este estudio se encuentran los siguientes: clasificación del texto de los tuits como spam, relación entre followers y followees y porcentaje de tuits que contienen enlaces. Finalmente, otra trabajo interesante es el de Ferrara et al. [40] en el que se estudian y clasifican varios enfoques para la detección de bots.

VIII. CONCLUSIONES Y TRABAJO FUTURO

En el estudio que presentamos en este trabajo se han evaluado 4 indicadores con el objetivo de detectar bots de Twitter. Estos 4 indicadores se han diseñado en base a 3 hipótesis sobre las diferencias entre las cuentas bot y las cuentas de usuarios humanos. Los indicadores 1 y 2, relativos al patrón de comportamiento temporal no se han mostrado efectivos por lo que la hipótesis **H1** no ha podido ser validada. En cambio, los resultados obtenidos para los indicadores 3 y 4 son prometedores y parecen validar las hipótesis **H2** y **H3**, relativas a la variabilidad del lenguaje empleado en las cuentas de Twitter y al porcentaje de tuits que contienen enlaces.

Como se ha explicado en la introducción, el estudio que se presenta en este trabajo es un primer paso que debe ser continuado en trabajos posteriores. A continuación se esbozan varias líneas de trabajo que planeamos desarrollar en el futuro. En primer lugar, es necesario ampliar el conjunto de indicadores que se utilizan. Dentro de este apartado se debe repensar la hipótesis **H1** y el uso de indicadores temporales que, como se ha indicado anteriormente no han dado los resultados esperados. Como se ha indicado en la sección de trabajos relacionados, articulos como [39] sugieren un patrón de comportamiento temporal de los bots en ráfagas.

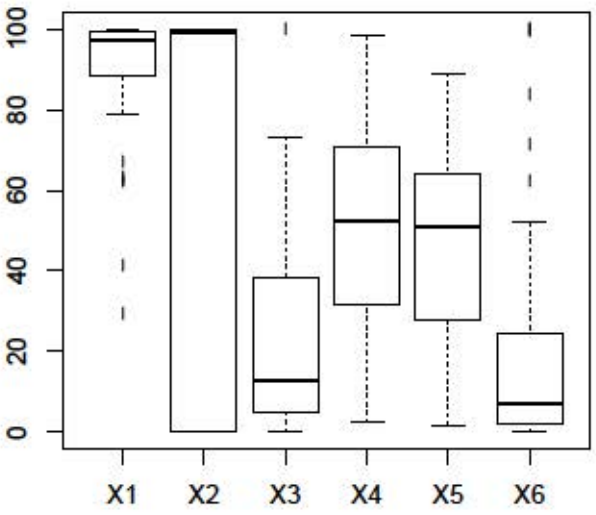

Fig. 5. Resultados para el indicador 4: Porcentaje de tuits que contienen enlaces

Se deben por lo tanto diseñar indicadores que detecten este tipo de comportamiento temporal. Por otra parte, el buen comportamiento observado por el indicador 3 nos anima a explorar el uso de otros indicadores lingüísticos, tanto relativos a la estructura sintáctica de las sentencias contenidas en una cuenta de Twitter como otros relativos al contenido de los tuits.

Una vez que se disponga de un conjunto de indicadores adecuado, será necesario entrenar un clasificador utilizando técnicas de aprendizaje máquina. Como parte de dicho proceso será necesario seleccionar los indicadores que se deben utilizar, para lo cual existen varios modelos estadísticos que se podrían considerar [41], [42]. Finalmente es necesario determinar cual de las múltiples técnicas de clasificación se va a utilizar (support vector machines, análisis discriminante de Fisher, modelo logit, Naive Bayes, etc.).

Finalmente, será necesario realizar una evaluación del clasificador así obtenido utilizando datos de un volumen de cuentas elevado (cientos de miles) para poder garantizar que los resultados obtenidos sean fiables.

La explotación en un entorno realista de una herramienta como la que se pretende desarrollar involucraría el procesado de grandes volúmenes de datos (Big Data) para aplicar estas técnicas a un gran número de cuentas de Twitter. Para tratar con esto sería interesante utilizar modelos predecibles de procesado distribuido de flujos, como el propuesto en [43].

AGRADECIMIENTOS

Este trabajo está parcialmente financiado por el Ministerio de Economía y Competitividad a través del proyecto "HERMES-SMARTDRIVER" (TIN2013-46801-C4-2-R) y por la Comunidad de Madrid a través del proyecto "eMadrid" (S2013/ICE-2715).

REFERENCIAS

[1] N. B. Ellison *et al.*, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[2] T. Berners-Lee, M. Fischetti, and M. L. Foreword By-Dertouzos, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper Information, 2000.

[3] B. A. Huberman and L. A. Adamic, "Internet: growth dynamics of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 131–131, 1999.

[4] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we rt?" in *first workshop on social media analytics*. ACM, 2010, pp. 71–79.

[5] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, "Collective attention in the age of (mis)information," *Computers in Human Behavior*, no. 0, pp. –, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563215000382

[6] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: collective narratives in the age of (mis) information," *PLoS ONE*, vol. 10, no. 2, 2015.

[7] M. L. Congosto, "Elecciones europeas 2014: Viralidad de los mensajes en twitter," *Redes: revista hispana para el analisis de redes sociales*, vol. 26, no. 1, pp. 23–52, 2015.

[8] M. L. Congosto, "Viralidad de los mensajes en twitter en las campañas electorales," in *III Congreso Internacional en Comunicación Política y Estrategias de Campana*, 2014.

[9] C. B. Williams and J. Girish, "Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008," *New Media & Society*, p. 1461444812457332, 2012.

[10] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, vol. 43, no. 3, pp. 345–354, 2006.

[11] C. C. Yang and T. D. Ng, "Terrorism and crime related weblog social network: Link, content analysis and information visualization," in *Intelligence and Security Informatics, 2007 IEEE*. IEEE, 2007, pp. 55–58.

[12] S. P. Borgatti and R. Cross, "A relational view of information seeking and learning in social networks," *Management science*, vol. 49, no. 4, pp. 432–445, 2003.

[13] D. Gayo-Avello, "Nepotistic relationships in twitter and their impact on rank prestige algorithms," *Information Processing & Management*, vol. 49, no. 6, pp. 1250–1280, 2013.

[14] F. Li and T. C. Du, "Who is talking? an ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs," *Decision Support Systems*, vol. 51, no. 1, pp. 190–197, 2011.

[15] S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication," *Journal of medical Internet research*, vol. 15, no. 4, 2013.

[16] S. A. Adams, "Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review," *International journal of medical informatics*, vol. 79, no. 6, pp. 391–400, 2010.

[17] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on twitter: human, bot, or cyborg?" in *Proceedings of the 26th annual computer security applications conference*. ACM, 2010, pp. 21–30.

[18] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[19] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.

[20] G. Tavares and A. Faisal, "Scaling-laws of human broadcast communication enable distinction between human, corporate and robot twitter users," *PloS one*, vol. 8, no. 7, p. e65774, 2013.

[21] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach (3rd edition)*. Prentice Hall, 2009.

[22] J. Weizenbaum, "Eliza-a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[23] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003, pp. 423–430.

[24] N. Bremmen, "The 100 most influential news media twitter accounts," http://memeburn.com/2010/09/the-100-most-influential-news-media-twitter-accounts/ (visitado el 7 de mayo de 2015), 2010.

[25] C. McCann, "The list: Ten fake twitter accounts," http://www.ft.com/cms/s/2/c843804c-3b21-11e2-b3f0-00144feabdc0.html#axzz2jz08haZW (visitado el 7 de mayo de 2015).
[26] I. Paul, "15 fake and funny twitter accounts," http://www.pcworld.com/article/159492/fake_funny_twitter.html (visitado el 7 de mayo de 2015), 2009.
[27] E. Levine, "The 15 best fake twitter accounts of 2012," http://heavy.com/comedy/2012/12/the-15-best-fake-twitter-accounts-of-2012/ (visitado el 7 de mayo de 2015), 2012.
[28] K. Knibbs, "The 8 best twitter bots you aren't following," http://www.digitaltrends.com/social-media/the-10-best-twitter-bots-you-arent-following/ (visitado el 7 de mayo de 2015), 2013.
[29] "The streaming apis," https://dev.twitter.com/streaming/overview (visitado el 7 de mayo de 2015), 2015.
[30] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate analysis*. Academic press, 1979.
[31] D. Peña, *Análisis de datos multivariantes*. Mc Graw Hill Interamericana de España, S.A.U., 2002.
[32] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 3, pp. 252–264, 1991.
[33] M. Lavine, "Introduction to statistical thought," http://people.math.umass.edu/~lavine/Book/book.html (visitado el 7 de mayo de 2015), 2005.
[34] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The fukushima disaster and information source credibility on twitter," in *9th International ISCRAM Conference*, 2012, pp. 1–10.
[35] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *26th Annual Computer Security Applications Conference*. ACM, 2010, pp. 1–9.
[36] A. H. Wang, "Don't follow me: Spam detection in twitter," in *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. IEEE, 2010, pp. 1–10.
[37] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," in *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, vol. 6, 2010, p. 12.
[38] A. H. Wang, "Detecting spam bots in online social networking sites: a machine learning approach," in *24th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy*. Springer, 2010, pp. 335–342.
[39] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 6, pp. 811–824, 2012.
[40] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *arXiv preprint arXiv:1407.5225*, 2014.
[41] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
[42] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*. Akademinai Kiado, 1973, pp. 267–281.
[43] P. Basanta, N. Fernandez-Garcia, A. Wellings, and N. Audsley, "Improving the predictability of distributed stream processors," *Future Generation Computing Systems*, 2015.