

definen diferencias bien demarcadas entre los distintos tipos de usuario.

Luego de estudiar el contenido de las publicaciones y el comportamiento de los usuarios (humanos, bots y ciborgs) en Twitter, se observó que existían una serie de características que podrían ser útiles para la clasificación automática de las distintas categorías de usuarios. Fueron estas características las utilizadas para el entrenamiento y evaluación de un modelo Random Forest que logró una precisión de un 86%, cuya debilidad se puede decir son los ciborgs, especulando que esto es debido a su naturaleza híbrida.

Cabe destacar que las tecnologías elegidas para el desarrollo del modelo (Apache Spark) hacen viable la evaluación de grandes cantidades de datos de forma distribuida y paralela, debido a los principios en los cuales está fundamentado.

REFERENCIAS

- [1] M. McGiboneyn, "Twitter's Tweet Smell Of Success", 2009. [En línea]. Disponible: <http://www.nielsen.com/us/en/insights/news/2009/twitters-tweet-smell-of-success.html>.
- [2] Twitter, Inc., "Empresa | About", 2016. [En línea]. Disponible: <https://about.twitter.com/es/company>.
- [3] E. Ferrara, O. Varol, C. Davis, F. Menczer, y A. Flammini, "The rise of social bots", en Communications of the ACM, vol. 59, iss. 7, Eds. ACM New York, pp. 96-104, 2016.
- [4] Y. Boshmaf, I. Muslukhov, K. Beznosov, y M. Ripeanu, "The socialbot network: when bots socialize for fame and money", en Proceedings of the 27th Annual Computer Security Applications Conference, pp. 93-102, 2011.
- [5] A. Murgia, D. Janssens, S. Demeyer, and B. Vasilescu, "Among the Machines: Human-Bot Interaction on Social Q&A Websites" en Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1272-1279, 2016.
- [6] G. Penna, "Reality Mining in Twitter", M.S.c degree, Dept. Comp., Imperial College London, London, GB, 2012.
- [7] Z. Chu, S. Gianvecchio, H. Wang, S. Jajosia, "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?," en IEEE Transactions on Dependable and Secure Computing, vol.9, no. 6, pp. 811-824, 2012.
- [8] R. Hanneman, "Introduction to social network methods", University of California, Riverside, CA, 2005. [En línea]. Disponible: <http://faculty.ucr.edu/~hanneman/>.
- [9] E. F. Moore. "The shortest path through a maze", en Int. Symp. on Th. of Switching, pp. 285-292, 1959.
- [10] Twitter, Inc., "API Rate Limits | Twitter Developers", 2016. [En línea]. Disponible: <https://dev.twitter.com/rest/public/rate-limiting>.
- [11] MINCI, "Venezuela amplía acceso a servicios de Internet", [En línea]. Disponible: <http://minci.gob.ve/2016/06/venezuela-amplia-acceso-a-servicios-de-internet/>.
- [12] Statista, "Number of active Twitter users in leading markets as of May 2016", 2016, [En línea]. Disponible: <http://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
- [13] S. Smith, "Determining Sample Size: How to Ensure You Get the Correct Sample Size", 2013, [En línea]. Disponible en: <https://www.qualtrics.com/blog/determining-sample-size/>.
- [14] T. K. Ho, "Random decision forests. In Document Analysis and Recognition", en Proceedings of the Third International Conference, vol. 1, pp. 278-282, 1995.
- [15] M. Mccord y M. Chuah, "Spam detection on twitter using traditional classifiers", en International Conference on Autonomic and Trusted Computing, pp. 175-186, 2011.
- [16] Z. S. Harris, "Distributional structure" Word, vol. 10, no. 2-3, pp. 146-162, 1954.
- [17] C. Gini, "Variabilità e mutabilità" Reimprimido en Pizetti, E.; Salvemini, T., eds. 1955. Memorie di metodologica statistica, Libreria Eredi Virgilio Veschi, Roma, IT , 1912.