



UNIVERSITÀ
DELLA CALABRIA

DIPARTIMENTO DI INGEGNERIA INFORMATICA,
MODELLISTICA, ELETTRONICA E SISTEMISTICA
(DIMES)

CORSO DI LAUREA MAGISTRALE IN
INGEGNERIA INFORMATICA

Analisi della polarizzazione politica e studio
dell'influenza dei bot durante le elezioni presidenziali
USA 2016

RELATORI

Ing. Fabrizio Marozzo

Ing. Riccardo Cantini

CANDIDATO

Francesco Pileggi

Matricola: 195309

A.A. 2019-2020

Indice

Capitolo 1	4
Introduzione	4
Capitolo 2	6
Big Data e Social Media.....	6
2.1 Social Big Data.....	8
2.2 Big Data.....	11
2.2.1 Sfide del mondo dei Big Data	13
2.2.2 Sorgenti di Big Data	14
2.2.3 Sistemi di Elaborazione.....	16
Capitolo 2	19
Bot Detection.....	19
2.1 Approcci di Bot Detection.....	21
2.1.1 Holistic Approach.....	21
2.1.2 Pairwaise Account Comparison	22
2.1.3 Egocentric Analysis.....	22
2.2 Tecniche di Bot Detection.....	23
2.2.1 DeBot	23
2.2.2 Twitter Sybils Detector (TSD)	29
2.2.3 Ensamble of Specialized Classifiers (ESC – Botometer v4.0).....	32
2.3 Botometer	36
2.3.1 Metodo Proposto	37
2.3.2 Feature Importance Analysis.....	39
2.3.3 Modello di validazione per Top N Features	40
Capitolo 3	43
Social Media Analysis per elezioni politiche	43
3.1 Tecniche di Social Media Analysis per opinione pubblica	45
3.2 Un algoritmo di Social Media Analysis: Marozzo-Bessi	46
3.2.1 Caratteristiche della metodologia	47
3.2.2 Descrizione della Metodologia.....	48
3.3 Metodologia proposta.....	51
3.3.1 Una variante della tecnica di Marozzo e Bessi.....	52
3.3.2 Una variante competitiva del Linear Treshold Model.....	55

Capitolo 4	57
Analisi della polarizzazione degli utenti social e dell'influenza dei bot	57
4.1 Caso di Studio	57
4.2 Dataset	59
4.3 Analisi dei risultati	61
4.3.1 Polarizzazione Utenti	61
4.3.2 Analisi Temporale	66
4.3.3 Modelli di pubblicazione dei contenuti	70
4.3.4 Flussi di mobilità	74
4.3.5 Grafo dell'influenza	77
Conclusioni.....	81
Bibliografia.....	84

Capitolo 1

Introduzione

I social media hanno avuto nell'ultimo decennio grande successo in tutto il mondo, sono stati e vengono tuttora utilizzati per condividere contenuti sia testuali che multimediali riguardanti ogni genere di discussione. Da circa 10 anni si è assistito alla formazione su tali piattaforme di ampi gruppi di discussione relativi ad importanti tematiche sociali o legate alla politica. I social media sono diventati dunque negli anni sorgenti di preziosi dati grazie ai quali è possibile studiare il comportamento e le abitudini degli utenti. Tuttavia questi potenti strumenti di comunicazione possono essere sfruttati per manipolare le discussioni online, cambiare la percezione pubblica delle entità politiche o persino tentare di influenzare le elezioni presidenziali. Recenti studi hanno dimostrato la presenza di entità guidate algoritmicamente che appaiono come utenti legittimi, ma che in realtà agiscono sulle reti social in maniera automatica per influenzare le discussioni di qualsiasi tipo, ma in gran parte politiche: tali sistemi, vengono definiti social bot. La diffusione su larga scala dell'utilizzo dei social, ed in linea generale del web, ha portato in dote quelli che vengono definiti come Big Data, ovvero collezioni di dati di grandi dimensioni, provenienti da fonti differenti, che possono essere generati continuamente e con un'elevata velocità. Il processo di analisi e di estrazione di conoscenza da tali dati, di supporto ai processi decisionali, prende il nome di Big Data Analytics, le cui applicazioni non hanno il solo scopo di estrarre informazioni per ricavarne un profitto diretto, ma possono essere anche usate in contesti come l'analisi dei comportamenti umani. L'obiettivo dell'elaborato è quello di analizzare il comportamento degli utenti che parteciparono alle discussioni online durante le elezioni presidenziali USA 2016. L'analisi mira a classificare le intenzioni di voto degli utenti calcolandone la polarizzazione a partire da un dataset contenente i tweet pubblicati da ognuno, mediante l'utilizzo di tecniche di Social Media Analysis per elezioni politiche, e ad analizzare

l'influenza che gli account bot hanno avuto sulla polarizzazione degli utenti stessi mediante tecniche di bot detection e modelli di diffusione.

Nel primo capitolo la discussione sarà incentrata sui Social Media, sulla loro diffusione e sull'importanza dei dati che generano, anche detti Big Social Data. Verranno quindi discusse le caratteristiche di quest'ultimi a partire dal mondo dei Big Data includendo possibili metodologie da adottare per poterli manipolare.

Nel secondo capitolo la discussione sarà incentrata sul dominio della Bot Detection. Verranno discussi i principali metodi di rilevamento degli account automatici sulle piattaforme di social media nonché le tecniche per poter classificare un account come automatico. Verranno illustrati i principali framework di successo concentrandosi particolarmente sul framework *Botometer*, quest'ultimo rappresenta la metodologia di *bot detection* utilizzata nell'elaborato.

Nel terzo capitolo verranno descritte le tecniche di Social Media Analysis per le elezioni politiche e verrà illustrata la metodologia proposta in questo elaborato. Verranno quindi discusse le tecniche utilizzate per stabilire la *polarizzazione* degli utenti e le loro caratteristiche, approfondendo la metodologia di *Marozzo e Bessi* utilizzata in fase di sperimentazione, della quale verrà proposta una variante. Inoltre verrà introdotto un algoritmo competitivo basato sul *Linear Threshold Model*, utilizzato per modellare la diffusione dello spread a partire dai seed set di bot delle fazioni contrapposte.

Nel quarto capitolo verrà applicata la metodologia proposta ad un caso reale, nel caso specifico e discussioni politiche relative alle elezioni presidenziali USA 2016 sulla piattaforma di microblogging Twitter, analizzando in maniera approfondita i risultati ottenuti.

Bibliografia

- [1] Loris Belcastro, Fabrizio Marozzo e Domenico Talia. "Programming models and systems for Big Data analysis". In: *International Journal of Parallel, Emergent and Distributed Systems* 34.6 (2019), pp. 632–652.
- [2] Linton C Freeman e Rosanna Memoli. *Lo sviluppo dell'analisi delle reti sociali: uno studio di sociologia della scienza*. Angeli, 2007.
- [3] Jinha Kim, Seung-Keol Kim e Hwanjo Yu. "Scalable and parallelizable processing of influence maximization for large-scale social networks?" In: *2013 IEEE 29th international conference on data engineering (ICDE)*. IEEE. 2013, pp. 266–277.
- [4] We Are Social Inc. *We are social*. 2008. url: <https://wearesocial.com/>
- [5] GlobalWebIndex Inc. *GlobalWebIndex*. 2009. url: <https://www.globalwebindex.com/>
- [6] Min Chen, Shiwen Mao e Yunhao Liu. "Big data: A survey". In: *Mobile networks and applications* 19.2 (2014), pp. 171–209.
- [7] Boshmaf, Yazan, et al. "Design and analysis of a social botnet." *Computer Networks* 57.2 (2013): 556-578.
- [8] Alsaleh, Mansour, et al. "Tsd: Detecting sybil accounts in twitter." *2014 13th International Conference on Machine Learning and Applications*. IEEE, 2014. [9] Chavoshi, Nikan, Hossein Hamooni, and Abdullah Mueen. "Debot: Twitter bot detection via warped correlation." *Icdm*. 2016. 2-7.
- [10] Sayyadilarikandeh, Mohsen, et al. "Detection of Novel Social Bots by Ensembles of Specialized Classifiers." *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020. 2-8.
- [11] Varol, Onur, et al. "Online human-bot interactions: Detection, estimation, and characterization." *arXiv preprint arXiv:1703.03107* (2017). 3-10.
- [12] Davis, Clayton Allen, et al. "Botornot: A system to evaluate social bots." *Proceedings of the 25th international conference companion on world wide web*. 2016. 1-6.
- [13] Yang, Kai-Cheng, et al. "Scalable and generalizable social bot detection through data selection." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 01. 2020. 2-9.
- [14] Dong, Guozhu, and Huan Liu, eds. *Feature engineering for machine learning and data analytics*. CRC Press, 2018. 312- 326

[15] Alarifi, Abdulrahman, Mansour Alsaleh, and AbdulMalik Al-Salman. "Twitter turing test: Identifying social machines." *Information Sciences* 372 (2016): 332-346.

[16] K. Wegrzyn-Wolska and L. Bougueroua, "Tweets mining for French Presidential Election," Proc. 4th Int'l Conf. on Computational Aspects of Social Networks (CASoN), Sao Carlos, 2012, pp. 138-143.

[17] T. Mahmood, T. Iqbal, F. Amin, W. Lohanna and A. Mustafa, "Mining Twitter big data to predict 2013 Pakistan election winner," INMIC, Lahore, 2013, pp. 49-54.

[18] J. M. Soler, F. Cuartero and M. Roblizo, "Twitter as a Tool for Predicting Elections Results," Proc. IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining, Istanbul, 2012, pp. 1194-1200.

[19] Marozzo Fabrizio, and Alessandro Bessi. "Analyzing polarization of social media users and news sites during political campaigns." *Social Network Analysis and Mining* 8.1 (2018): 1.

[20] L. Belcastro, R. Cantini, F. Marozzo, D. Talia and P. Trunfio, "Learning Political Polarization on Social Media Using Neural Networks," in *IEEE Access*, vol. 8, pp. 47177-47187, 2020, doi: 10.1109/ACCESS.2020.2978950.

[21] K. Jaidka, S. Ahmed, M. Skoric, and M. Hilbert, "Predicting elections from social media: a three-country, three-method comparative study," *Asian Journal of Communication*, vol. 29, no. 3, pp. 252–273, 2019.

[22] M. Bilal, A. Gani, M. Marjani, and N. Malik, "Predicting elections: Social media data and techniques," in 2019 International Conference on Engineering and Emerging Technologies (ICEET). IEEE, 2019, pp. 1–6.

[23] Gaurav, Manish, et al. "Leveraging candidate popularity on Twitter to predict election outcome." *Proceedings of the 7th workshop on social network mining and analysis*. 2013.

[24] Tumasjan, Andranik, et al. "Predicting elections with twitter: What 140 characters reveal about political sentiment." *Fourth international AAAI conference on weblogs and social media*. 2010.

[25] Burnap, Pete, et al. "140 characters to victory?: Using Twitter to predict the UK 2015 General Election." *Electoral Studies* 41 (2016): 230-233.

[26] Hafner-Burton, Emilie M., and Alexander H. Montgomery. "Centrality in politics: How networks confer power." (2010).

[27] Miller, Patrick R., et al. "Talking politics on Facebook: Network centrality and political discussion practices in social media." *Political Research Quarterly* 68.2 (2015): 377-391.

[28] D.Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, pages 137–146, New York, NY, USA, 2003.

[29] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Who is tweeting on Twitter: Human, bot, or cyborg? In Proc. of the Annual Conf. on Computer Security Applications, pages 21–30, 2010.

[30] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? IEEE Tran Dependable & Secure Comput, 9(6):811–824, 2012.

[31] Eric Clark, Chris Jones, Jake Williams, Allison Kurti, Michell Nortotsky, Christopher Danforth, and Peter Dodds. Vaporous marketing: Uncoverin, pervasive electronic cigarette advertisements on Twitter. Preprint 1508.01843, arXiv, 2015.