You have been CAUTE! Early Detection of Compromised Accounts on Social Media

By: VanDam, C (VanDam, Courtland)[1,2]; Masrour, F (Masrour, Farzan)[2]; Tan, PN (Tan, Pang-Ning)[2]; Wilson, T (Wilson, Tyler)[2]

PROCEEDINGS OF THE 2019 IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM 2019)
Edited by: Spezzano, F; Chen, W; Xiao, X
Pages: 25-32
DOI: 10.1145/3341161.3342868
Published: 2019
Document Type: Proceedings Paper

Conference

Conference: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)
Location: Vancouver, CANADA
Date: AUG 27-30, 2019
Sponsor(s): IEEE; Assoc Comp Machinery; IEEE Comp Soc; ACM SIGKDD; IEEE TCDE; Springer; Elsevier

Abstract

Detection of compromised social media accounts is an important problem as the compromised accounts can be exploited by hackers to spread false and misleading information. In particular, early detection of compromised accounts is essential to mitigating the damages caused by the hackers' posts, which may range from victim shaming to causing widespread public panic and civil unrest. This paper proposes CAUTE, a deep learning framework that simultaneously learns the feature embeddings of the users and their posts in order to identify which, if any, of their posts were written by a different person, i.e. a hacker. Using Twitter as an example of the social media platform, CAUTE learns a tweet-to-user encoder to infer the user features from tweet features and a user-to-tweet encoder to predict the tweet content from a combination of the user features and the tweet meta features. The residual errors of both encoders are then fed into a fully-connected neural network layer to detect whether a post was published by the specified user or by a hacker. Experimental results showed that the features learned by CAUTE are more informative than those generated by conventional representation learning methods. Additionally, CAUTE outperformed several state-of-the-art baseline algorithms in terms of their overall performance and can effectively detect compromised posts early without generating too many false alarms.

Author Information

Reprint Address:
Lincoln Laboratory Massachusetts Institute of Technology (MIT) MIT, Lincoln Lab, 244 Wood St, Lexington, MA 02173 USA.
Corresponding Address: VanDam, C (corresponding author)

MIT, Lincoln Lab, 244 Wood St, Lexington, MA 02173 USA.
Organization-Enhanced Name(s)
Lincoln Laboratory
Massachusetts Institute of Technology (MIT)

Addresses:

[1] MIT, Lincoln Lab, 244 Wood St, Lexington, MA 02173 USA
Organization-Enhanced Name(s)
Lincoln Laboratory
Massachusetts Institute of Technology (MIT)

[2] Michigan State Univ, E Lansing, MI 48824 USA
Organization-Enhanced Name(s)
Michigan State University

E-mail Addresses: courtland.vandam@ll.mit.edu; masrour@cse.msu.edu; ptan@cse.msu.edu; wils1270@msu.edu

Funding

Funding Agency Show details	Grant Number
National Science Foundation (NSF)	IIS-1615612

Close funding text
This work is supported in part by the U.S. National Science Foundation under grant #IIS-1615612.

Publisher

ASSOC COMPUTING MACHINERY, 1515 BROADWAY, NEW YORK, NY 10036-9998 USA

Categories / Classification

Research Areas: Computer Science; Telecommunications

Web of Science Categories: Computer Science, Artificial Intelligence; Computer Science, Information Systems; Telecommunications

Document Information

Language: English

Accession Number: WOS:000555683800004

ISBN: 978-1-4503-6868-1

Other Information

IDS Number: BP5IK

Cited References in Web of Science Core Collection: 16

Times Cited in Web of Science Core Collection: 0

See fewer data fields

Cited References: 16

Showing 16 of 16 **View All in Cited References page** *(from Web of Science Core Collection)*

1. **Authorship verification applied to detection of compromised accounts on online social networks** **Times Cited: 2**
By: Barbon, S.; Igawa, R. A.; Zarpelao, B. B.
Multimedia Tools and Applications Volume: 76 Issue: 3 Published: 2017

2. **Detecting Compromised Accounts on the Pokec Online Social Network** **Times Cited: 3**
By: Bohacik, Jan; Fuchs, Antonin; Benedikovic, Miroslav
2017 INTERNATIONAL CONFERENCE ON INFORMATION AND DIGITAL TECHNOLOGIES (IDT) Pages: 56-60 Published: 2017

3. **COMPA: Detecting Compromised Accounts on Social Networks** **Times Cited: 9**
By: Egele, M.; Stringhini, G.; Kruegel, C.; et al.
ISOC NETW DISTR SYST Published: 2013
Publisher: Internet Society, San Diego, California
[Show additional data]

4. **Towards Detecting Compromised Accounts on Social Networks** **Times Cited: 26**
By: Egele, Manuel; Stringhini, Gianluca; Kruegel, Christopher; et al.
IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING Volume: 14 Issue: 4 Pages: 447-460 Published: JUL-AUG 2017

5. **Recognition of compromised accounts on twitter** Times Cited: 1
By: Igawa, R. A.; de Almeida, A. M. G.; Zarpelao, B. B.; et al.
SBSI 2015 Volume: 1 Published: 2015
Publisher: ACM, New York, NY, USA
[Show additional data]

6. **End-to-end compromised account detection** Times Cited: 1
By: Karimi, H.; VanDam, C.; Ye, L.; et al.
ASONAM 2018 Published: 2018
[Show additional data]

7. **Distributed representations of sentences and documents** Times Cited: 3
By: Le, Q.; Mikolov, T.
- Volume Volume: 32 Published: 2014

8. **Detecting Hacked Twitter Accounts by Examining Behavioural Change using Twitter Metadata** Times Cited: 2
By: Nauta, M.
P 25 TWENT STUD C IT Published: 2016

9. **Americans and cybersecurity** Times Cited: 3
By: Olmstead, K.; Smith, A.
Tech. Rep. Published: 2017
Online
Publisher: Pew Res. Center, Washington, DC, USA
URL: https://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/

10. **False ap twitter message sparks stock-market selloff** Times Cited: 17
By: Ovide, S.
Wall Street Journal Volume: 4 Published: 2013
Online Available
URL: https://www.wsj.com/articles/SB10001424127887323735604578440971574897016

11. **Profiling Online Social Behaviors for Compromised Account Detection** Times Cited: 33
By: Ruan, Xin; Wu, Zhenyu; Wang, Haining; et al.
IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY Volume: 11 Issue: 1 Pages: 176-187 Published: JAN 2016

12. **Consequences of connectivity: Characterizing account hijacking on twitter** Times Cited: 1
By: Thomas, K.; Li, F.; Grier, C.; et al.
CCS 14 Published: 2014
[Show additional data]

13. **Evaluating Algorithms for Detection of Compromised Social Media User Accounts** Times Cited: 8
By: Trang, David; Johansson, Fredrik; Rosell, Magnus
SECOND EUROPEAN NETWORK INTELLIGENCE CONFERENCE (ENIC 2015) Pages: 75-82 Published: 2015

14. **Cadet: Compromised account detection using unsupervised learning** Times Cited: 1
By: VanDam, C.; Tan, P.-N.; Tang, J.; et al.
ASONAM 2018 Published: 2018
2018
[Show additional data]

15. **Understanding compromised accounts on twitter** Times Cited: 1
By: VanDam, C.; Tang, J.; Tan, P.-N.
WI 17 Published: 2017

16. **Towards detecting anomalous user behavior in online social networks** Times Cited: 2
By: Viswanath, B.; Bashir, M. A.; Crovella, M.; et al.
23 USENIX SEC S USEN Published: 2014
[Show additional data]

Showing 16 of 16 **View All in Cited References page**