Social Media Anomaly Detection: Challenges and Solutions

Yan Liu
University of Southern California
941 Bloom Walk
Los Angeles, CA 90292
yanliu.cs@usc.edu

Sanjay Chawla
Qatar Computing Research Institute and
University of Sydney
schawla@qf.org.qa

ABSTRACT

Anomaly detection is of critical importance to prevent malicious activities such as bullying, terrorist attack planning, and fraud information dissemination. With the recent popularity of social media, new types of anomalous behaviors arise, causing concerns from various parties. While a large body of work have been dedicated to traditional anomaly detection problems, we observe a surge of research interests in the new realm of social media anomaly detection. In this tutorial, we survey existing work on social media anomaly detection, focusing on the new anomalous phenomena in social media and most recent techniques to detect those special types of anomalies. We aim to provide a general overview of the problem domain, common formulations, existing methodologies and future directions.

Keywords

Anomaly detection; Social media analysis

1. INTRODUCTION

Social media systems provide convenient platforms for people to share, communicate, and collaborate. Anomalous behavior in social media may give early indication of emerging events, such as natural disasters and disease outbreak, such that people can be better informed and prepared for these events. Meanwhile, many malicious behaviors, such as bullying, terrorist attack planning, and fraudulent information dissemination, can take advantage of social media to induce worse outcome. Therefore it is extremely important that we can detect these abnormal activities as accurately and early as possible to prevent disasters and attacks. Needless to say, as more social information becomes available, the most challenging question is what useful patterns could be extracted from this influx of time-series social media data to help with the detection task.

Different from traditional anomaly detection scenarios, social media provides us with additional social information

between users (e.g. social networks between users). We can summarize the anomalies into two types: one is point anomaly, which refers to the abnormal behaviors of individual users. The other type of anomaly is group anomaly, which involves hidden groups of people (also known as tribe) working collectively to achieve a malicious goal. Another way to summarize the anomalies is by their input type, which usually consists of sequential data, unstructured texts, and graphs.

This tutorial aims to provide the participants a broad and comprehensive coverage on the foundations and recent developments on anomaly detection social media. We will provide both theoretical and practical results as well as illustrative demos. In contrast to previous tutorials on anomaly detection, we will focus on presenting and discussing a broad coverage of the emerging approaches of anomaly detection algorithms in the context of scalability and practicability. We will also offer useful and complementary information to the WSDM community for whom prepare for pursuing this research area.

2. OUTLINE OF THE TUTORIAL

The 3-hour tutorial consists of two parts and a break in the middle. In Part 1, we will provide an introduction to anomaly detection for social media data, including an overview of anomaly detection, data types and properties of social media data, anomaly detection in network data [1, 2, 7, 13], and anomaly detection in temporal data [3, 4, 8, 9]. In Part 2, we will discuss recent advances in anomaly detection for social media data, including point anomaly detection in social media [6, 10], group anomaly detection in social media [5, 11], as well as applications and systems.

To summarize, we will

- Present a balanced review of the area of anomaly detection for social media data by presenting topics of both practical and theoretical interest
- Describe the state-of-the-art and emerging analysis technologies on massive social media data in order to identify the recent and future trends
- Provide a good starting point, including tutorial slides, supplementary survey paper, implementation packages and data repository with real application datasets, for researchers entering this active research area by looking at both system- and algorithmic-level developments.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM 2017 February 06-10, 2017, Cambridge, United Kingdom

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4675-7/17/02.

DOI: <http://dx.doi.org/10.1145/3018661.3022757>

3. PREVIOUS OFFERING OF THE TUTORIAL

This tutorial was offered in ACM SIGKDD Conference in 2015. The slides will be available beforehand on our tutorial website. A sample of the slides can be accessed through the KDD 2015 tutorial website: <http://www-bcf.usc.edu/~liu32/SMAD.htm>. We also have a survey paper on social media anomaly detection [12], in which the audience can find more details of the algorithms discussed in the tutorial.

4. ACKNOWLEDGMENTS

Yan Liu's research was partly sponsored by the NSF research grants IIS-1134990 and IIS-1254206, and U.S. Defense Advanced Research Projects Agency (DARPA) under Anomaly Detection at Multiple Scales (ADAMS) program and Social Media in Strategic Communication (SMISC) program. Sanjay Chawla's research is partially supported by the Australian Research Council, Capial Markets CRC, CRC for Alertness, Sleep and Productivity. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agency, or the U.S. Government.

5. REFERENCES

- [1] L. Akoglu and M. McGlohon. Anomaly detection in large graphs. *CMU-CS-09-173 Technical*, (November), 2009.
- [2] L. Akoglu, M. McGlohon, and C. Faloutsos. Oddball: Spotting anomalies in weighted graphs. In *Advances in Knowledge Discovery and Data Mining*, pages 410-421. Springer, 2010.
- [3] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection for discrete sequences: A survey. *IEEE Trans. on Knowl. and Data Eng.*, 24(5):823-839, May 2012.
- [4] H. Cheng, P.-N. Tan, C. Potter, and S. Klooster. Detection and characterization of anomalies in multivariate time series. In *Proceedings of SDM '09*, pages 413-424, 2009.
- [5] L. Friedland and D. Jensen. Finding tribes: identifying close-knit individuals from employment patterns. In *Proceedings of KDD '07*, pages 290-299, New York, NY, USA, 2007. ACM.
- [6] Y. Ge, H. Xiong, C. Liu, and Z.H. Zhou. A Taxi Driving Fraud Detection System. In *Proceedings of ICDM '11*, pages 181 - 190, 2011.
- [7] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In *Proceedings of ICDM '05*, pages 418 - 425, Washington, DC, USA, 2005. IEEE Computer Society.
- [8] H. Qiu, Y. Liu, N. Subrahmanya, W. Li. Granger Graphical Models for Time-Series Anomaly Detection. In *Proceedings of ICDM' 2012*, Pages 1074-1079, 2012.
- [9] P. Sun, S. Chawla, and B. Arunasalam. Mining for outliers in sequential databases. In *Proceedings of SIAM '06*, pages 94-105, 2006.
- [10] W.-K. Wong, A. Moore, G. Cooper, and M. Wagner. Bayesian network anomaly pattern detection for disease outbreaks. In *Proceedings of ICML '03*, 2003.
- [11] Q. Yu, X. He, and Y. Liu. GLAD: Group anomaly detection in social media analysis. In *Proceedings of KDD '14*, Pages 372-381, 2014.
- [12] Rose Yu, Huida Qiu, Zhen Wen, ChingYung Lin, and Yan Liu. A Survey on Social Media Anomaly Detection. *SIGKDD Explor. Newsl.* 18, 1 (August 2016), 1-14.
- [13] H. Tong, C.-Y. Lin. Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection. In *Proceedings of SDM '2011*, pages 143-153, 2011.