

Collective Sensemaking via Social Sensors: Extracting, Profiling, Analyzing, and Predicting Real-world Events

Yuheng Hu
Department of Information and
Decision Sciences
University of Illinois at Chicago
yuhenghu@uic.edu

Yu-Ru Lin
School of Information
Sciences
University of Pittsburgh
yurulin@pitt.edu

Jiebo Luo
Department of Computer
Science
University of Rochester
jluo@cs.rochester.edu

ABSTRACT

Social media platforms like Twitter and Facebook have emerged as some of the most important platforms for people to discover, report, share, and communicate with others about various public events, be they of global or local interest (some high profile examples include the U.S Presidential debates, the Boston bombings, the hurricane Sandy, etc). The burst of social media reaction can be seen as a valuable real-time reflection of events as they happen, and can be used for a variety of applications such as computational journalism. Until now, such analysis has been mostly done manually or through primitive tools. Scalable and automated approaches are needed given the massive amounts of both event and reaction information. These approaches must also be able to conduct in-depth analysis of complex interactions between an event and its audience. Supporting such automation and examination however poses several computational challenges. In recent years, research communities have witnessed a growing interest in tackling these challenges. Furthermore, much recent research has begun to focus on solving more complex event analytics tasks such as post-event effect quantification and event progress prediction. This tutorial aims to review and examine current state of the research progress on this emerging topic.

Keywords

Event analysis; Social media

1. INTRODUCTION

Increasingly rich and large-scale social multimedia data (including text, images, audio, video) are being generated and posted to social media channels. Real-time social media streams have many advantages over the traditional media channels, such as ubiquity, mobility, immediacy, and seamless communication in reporting, covering and sharing real-world events (e.g, SuperBowl, State of the Union). Given these advantages, social media posts such as tweets can typically reflect events as they happen in real-time. Despite its

benefits, social media also tends to be noisy, chaotic, and overwhelming. As a result, the vast amount of noisy social media posts poses tremendous challenges for conducting automated in-depth tweets examination, which is critical to applications for journalistic investigation, playback of events, storytelling, etc.

In this tutorial we will examine the current state-of-the-art in supporting such automation and examination on the crowd's social media posts in the context of the public events that they are in response to. We will begin with an overview of recent technical innovations from the Machine learning, Data mining, Multimedia, Computer Vision and Social computing communities in handling several core research tasks – namely, detection and extraction of events from social multimedia, quantification and prediction of event impact, sensing and characterization of user activities, and estimation of user profiles.

Throughout the tutorial, we will discuss about the advantages and disadvantages of various approaches. At the end, we will also highlight the current open challenges in handling more complex event analytics tasks using social multimedia data.

2. TUTORIAL DESCRIPTION

2.1 Target Audience

Researchers in the field of analysis of user-generated content will benefit the most as this will give them an exhaustive overview of the research in the direction of event analysis using social multimedia data. We believe that the tutorial will give the newcomers a complete picture of the current work, introduce important research topics in this field, and inspire them to learn more. People from industry will benefit from the discussions on a large number of applications where such mechanisms have already been applied.

2.2 Goal of Tutorial

A summary of the current state of the research progress in computational methods to extract, model, analyze and predict public events using social multimedia data. The intent of the tutorial is to provide an overview of the current research and challenges in the area. It will survey and explore the recent advances in Data Mining, Social Computing, Machine Learning, Computer Vision, and Multimedia communities involving various supervised and unsupervised approaches. Particular attention will be given to examining and comparing the advantages and disadvantages of these approaches. In this context, future research directions in

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).

KDD '16 August 13-17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4232-2/16/08.

DOI: <http://dx.doi.org/10.1145/2939672.2945384>

event analytics and predictions for events from different domains will also be discussed.

2.3 Tutorial Content

Proposed duration of the tutorial is *three* hours. Below is the outline:

1. **Introduction and motivation.** We start the tutorial with real examples showing the usage of social media in covering major events of different types (e.g., Hurricane Sandy, Presidential debate, World cup). We show the need of automation and examination and computational challenges [14].
2. **Detect and extract events from social media.** We proceed by describing event detection algorithms in both supervised and Unsupervised fashion [11, 1]. We also discuss event extraction algorithms [12]. We show application of Event detection and extraction over social media for sport events, local festivals, Drug related adverse events, earthquakes, etc [13, 10].
3. **Quantify event's effect on its audience.** We proceed by showing the important of aligning event with social media posts and discuss several data mining approaches for the task [5, 6]. We also discuss how to find credibility and dynamics of events on social media [2, 9, 8, 4, 7].
4. **Sensing users and social activities from social multimedia.** We discuss the current status of social multimedia, key elements for understanding users in social multimedia, and accurate and comprehensive estimation of user profile and context. We also describe algorithms and applications based on mining social activities from social multimedia, such as suggestion of social groups from user's personal photo collection, and forecasting of election outcome based on image sharing activities and image sentiments [3, 16, 15].
5. **Summary and a look ahead.** We end the tutorial with a brief summary and identify open issues and research questions for sensemaking of events using social media in near future.

3. PRESENTER INFORMATION

Yuheng Hu is an assistant professor at College of Business Administration, University of Illinois at Chicago. Yuheng works at the interface of Data Mining, Social Computing, HCI and Machine Learning. His research focuses on developing algorithms, tools and systems to characterize, make sense of, and predict people's reactions on social media in response to different real world events. His work has been published at various highly reputed venues including AAAI, IJCAI, ICDE, TKDE, ICWSM, CHI, where he won a best paper nomination in 2013. His work has also been featured in press outlets such as ABC, PBS, The Seattle Times, and FastCompany.

Yu-Ru Lin is an assistant professor at the School of Information Sciences, University of Pittsburgh. Her research interests include human mobility, social and political network dynamics, and computational social science. Her current research focuses on extracting system-level features from big

data sets, including social media data and anonymized cell-phone records, for studying human and social dynamics, particularly under exogenous events such as emergencies and media events. Her work has appeared in prestigious scientific venues including WWW, SIGKDD, InfoVis, ACM TKDD, ACM TOMCCAP, IEEEE and PLoS ONE.

Jiebo Luo is a professor at the computer science department at the University of Rochester. He has been involved in numerous technical conferences, including serving as the program co-chair of ACM Multimedia 2010 and IEEE CVPR 2012. He is the Editor-in-Chief of the Journal of Multimedia, and has served on the editorial boards of prestigious journals. He has authored over 270 technical papers and 90 US patents. Prof. Luo is a Fellow of the SPIE, IEEE, and IAPR, as well as a member of the ACM and AAAI. His research spans image processing, computer vision, machine learning, data mining, medical imaging, and ubiquitous computing. A recent research thrust focuses on exploiting social media for machine learning, data mining, and human-computer interaction, for example, mining the wisdom of crowds for social, political, and economic prediction and forecasting.

4. REFERENCES

- [1] H. Becker, M. Naaman, and L. Gravano. Beyond trending topics: Real-world event identification on twitter. 2011.
- [2] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *WWW*. ACM, 2011.
- [3] L. Duan, D. Xu, I.-H. Tsang, and J. Luo. Visual event recognition in videos by learning from web data. *PAMI*, 2012.
- [4] Y. Hu, S. D. Farnham, and A. Monroy-Hernández. Whooly: facilitating information seeking for hyperlocal communities using social media. In *CHI*. ACM, 2013.
- [5] Y. Hu, A. John, D. D. Seligmann, and F. Wang. What were the tweets about? topical associations between public events and twitter feeds. In *ICWSM*, 2012.
- [6] Y. Hu, A. John, F. Wang, and S. Kambhampati. Et-lda: Joint topic modeling for aligning events and their twitter feedback. In *AAAI*, 2012.
- [7] Y. Hu, F. Wang, and S. Kambhampati. Listening to the crowd: Automated analysis of events via aggregated twitter sentiment. In *IJCAI*, 2013.
- [8] Y.-R. Lin, D. Margolin, B. Keegan, A. Baronchelli, and D. Lazer. # bigbirds never die: Understanding social dynamics of emergent hashtag. 2013.
- [9] Y.-R. Lin, D. Margolin, B. Keegan, and D. Lazer. Voices of victory: A computational focus group framework for tracking opinion shift in real time. In *WWW*, 2013.
- [10] J. Nichols, J. Mahmud, and C. Drews. Summarizing sporting events using twitter. In *IUI*, 2012.
- [11] A.-M. Popescu, M. Pennacchiotti, and D. Paranjpe. Extracting events and event descriptions from twitter. In *WWW*. ACM, 2011.
- [12] A. Ritter, O. Etzioni, S. Clark, et al. Open domain event extraction from twitter. In *KDD*, 2012.
- [13] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW*, 2010.
- [14] D. Shamma, L. Kennedy, and E. Churchill. Tweetgeist: Can the twitter timeline reveal the structure of broadcast events. *CSCW*, 2010.
- [15] Q. You and J. Luo. Towards social imagematics: sentiment analysis in social multimedia. In *Proceedings of the Thirteenth International Workshop on Multimedia Data Mining*, 2013.
- [16] J. Yuan, Q. You, and J. Luo. Are there cultural differences in event driven information propagation over social media? In *Proceedings of the 2nd international workshop on Socially-aware multimedia*, 2013.