Tutorial Proposal for SBP 2016
Exploring Public Events Using Social Media Data: Challenges in Extraction, Modeling and Analyzing

Yuheng Hu
University of Illinois at Chicago

I. Goal of Tutorial
A. Target Audience:

Researchers and graduate students with a background in social computing, data mining, natural language processing and machine learning.

B. Audience will walk away with:

A summary of the current state of the research progress in computational methods to extract, model, and analyze public events using social media 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 social computing, NLP and communities involving various supervised and unsupervised approaches to detect and extract public events from social media data such as Twitter posts (tweets), model events' topical and sentiment evolutions using corresponding tweets, and quantify the events' effects on Twitter audience. Particular attention will be given to examining and comparing the advantages and disadvantages of these approaches. In this context, future research directions in event analytics and predictions for events from different domains will also be discussed.

C. Innovative Aspects

Social media channels 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. Just to name a few:

- **Computational journalism:** *Journalists can leverage event-related social media responses to conduct news information gathering, organization, sensemaking, investigation and dissemination.*
- **Opinion/Sentiment mining:** *Past research has shown that massive number of people, companies and political campaigns turn to social media to collect views about products, services, political candidates during and after the event, etc. Gleaning insights from event-related responses can guide people's choices, decision-making, voting, and even stock market investments.*

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, the social computing, NLP and machine learning communities have witnessed a growing interest in tackling these challenges. Such new interest is evidenced by the steady growth of relevant papers appearing in various conferences such as AAAI, IJCAI, KDD, WWW, ACL, ICWSM. 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 will examine current state of the research progress on this emerging topic. In particular, it will focus on automated solutions to address three core research tasks in event analytics using unstructured tweets. To the best of our knowledge, this will be the first tutorial on event analytics given at any AI, NLP or social computing conference.**QN: ARE YOU SURE NOTHING ON EVENT ANALYTICS EVEN IN KDD/Datamining communities?***

II. Estimated audience size
We expect roughly 50-100 people to attend our tutorial.

III. Content

I. Introduction and motivation [30 min]
- Real examples showing the usage of social media in covering major events of different types
    a. Hurricane Sandy [], 2012 Presidential debate [],World cup 2010 []
- The need of automation and examination and computational challenges
    a. Current practices: manual approaches, word cloud, Tfidf, etc
    b. The need for automated approaches to conduct in-depth examination, and the challenges in supporting automation and examination

II. Detect and extract events from Twitter data [45 min]
- Infer event types: classification of planned vs. unplanned events []
- Event detection
    a. Supervised approaches: classification based approaches, hashtag based
    b. Unsupervised approaches: time-series based, clustering based
- Event extraction
    a. Open domain information extraction for events [], extract event descriptions and context []

III.   Characterize event with tweets [60 min]
- Align event with tweets
    a. time-window based [], generative models []
- Infer event's sentiment from tweets
    a. Sentiment analysis for tweets

IV.   Quantify event's effect on its audience [60 min]
- Dynamics of event related hashtag []
- Opinion shift w.r.t event progress []

V. Conclusion and future research directions [30 min]

IV. Tutorial Description

Social media channels enjoy many advantages over the traditional media channels, such as ubiquity, mobility, immediacy, and seamless communication in reporting, covering and sharing real-world public events. 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 tweets 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 tweets in the context of the public events that they are in response to. We will focus on details of recent technical innovations from the machine learning, NLP and social computing communities in handling three core research tasks. Namely, event detection and extraction from tweets, event characterization using tweets, and event effect quantification. Specifically, we will cover supervised and unsupervised approaches to detect public trending events and extract the event descriptions from noisy unstructured tweets streams. After that, we will cover approaches to characterize the events in terms of their topical and sentiment evolution. Next, we will present various frameworks to sample, monitor and quantify the event audience's behavior as well as its social network structures with respect to the progress of the event. 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 media data such as event progress prediction.

V. Prerequisite Knowledge

Basic knowledge of concepts in machine learning and social network analysis

VI. Presenter's Background

(List of publications in the area in are in references.)

Yuheng Hu
Dept. of Information and Decision Sciences.,
College of Business Administration
University of Illinois at Chicago
Email: yuhenghu@uic.edu

Yuheng Hu is an assistant professor at 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 conferences including AAAI, IJCAI, ICWSM and 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.

References

1. Lin, Y.-R., Margolin, D., Keegan, B., Baronchelli, A. and Lazer, D. (2013). #Bigbirds Never Die: Understanding Social Dynamics of Emergent Hashtag, ICWSM 2013
2. Lin, Y.-R., Margolin, D., Keegan, B. and Lazer, D. (2013). Voices of Victory: A Computational Focus Group Framework for Tracking Opinion Shift in Real Time, WWW 2013
3. H Becker, M Naaman, L Gravano, Beyond Trending Topics: Real-World Event Identification on Twitter. ICWSM, 2011
4. T Sakaki, M Okazaki, Y Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors. WWW 2010
5. AL Hughes, L Palen, Twitter adoption and use in mass convergence and emergency events - International Journal of Emergency Management, 2009
6. AM Popescu, M Pennacchiotti, D, Extracting events and event descriptions from twitter Paranjpe ICWSM 2011
7. D Shamma, L Kennedy, E Churchill, Tweetgeist: Can the twitter timeline reveal the structure of broadcast events, CSCW 2010
8. H Becker, M Naaman, L Gravano, Selecting Quality Twitter Content for Events. ICWSM, 2011
9. S Petrovic, M Osborne, R McCreadie, C Macdonald, Can Twitter replace Newswire for breaking news? ICWSM, 2013
10. Yuheng Hu, Ajita John, Doree Duncan Seligmann, Fei Wang, What were the Tweets about? Topical Associations between Public Events and Twitter Feeds. ICWSM'12
11. Yuheng Hu, Ajita John, Fei Wang, Subbarao Kambhampati, ET-LDA: Joint Topic Modeling for Aligning Events and their Twitter Feedback. AAAI 2012
12. Yuheng Hu, Fei Wang, Subbarao Kambhampati. Listen to the Crowd: Automated Analysis of Events via Aggregated Twitter Sentiment. IJCAI 2012
13. J Nichols, J Mahmud, C Drews, Summarizing sporting events using twitter. IUI2012
14. Ruihong Huang and Ellen Riloff. Modeling Textual Cohesion for Event Extraction. AAAI 2012