Towards a Social Media Analytics Platform: Event Detection and User Profiling for Twitter

— A Tutorial at WWW 2014 —

Manish Gupta
Microsoft, India
gmanish@microsoft.com

Rui Li
Yahoo! Inc.
ruililab@yahoo-inc.com

Kevin Chen-Chuan Chang
Univ. of Illinois at Urbana-Champaign
kcchang@illinois.edu

1. Abstract

Microblog data differs significantly from the traditional text data with respect to a variety of dimensions. Microblog data contains short documents, SMS kind of language, and is full of code mixing. Though a lot of it is mere social babble, it also contains fresh news coming from human sensors at a humungous rate. Given such interesting characteristics, the world wide web community has witnessed a large number of research tasks for microblogging platforms recently. Event detection on Twitter is one of the most popular such tasks with a large number of applications. The proposed tutorial on social analytics for Twitter will contain three parts. In the first part, we will discuss research efforts towards detection of events from Twitter using both the tweet content as well as other external sources. We will also discuss various applications for which event detection mechanisms have been put to use. Merely detecting events is not enough. Applications require that the detector must be able to provide a good description of the event as well. In the second part, we will focus on describing events using the best phrase, event type, event timespan, and credibility. In the third part, we will discuss user profiling for Twitter with a special focus on user location prediction. We will conclude with a summary and thoughts on future directions.

2. Presenters

- Manish Gupta is an applied researcher at the Bing team in Microsoft India R&D Private Limited at Hyderabad, India. He is also a visiting faculty at International Institute of Information Technology, Hyderabad. He received his Masters in Computer Science from IIT Bombay in 2007 and his Ph.D. from the University of Illinois at Urbana-Champaign in 2013. He worked for Yahoo! Bangalore for two years. His research interests are in the areas of web mining, data mining and information retrieval. He has published more than 25 research papers in referred journals and conferences, including ICDE, KDD, PKDD, SDM, WWW conferences.

- Rui Li is a scientist at Yahoo! lab. Before joining Yahoo! lab, he obtained his PhD in the Data and Information Systems (DAIS) Lab at UIUC in 2013. He was advised by Professor Kevin Chen-Chuan Chang. He received his Bachelor’s degree in Computer Science ACM Honor Class from Shanghai Jiaotong University, China in 2007.

- Kevin Chen-Chuan Chang is an Associate Professor in Computer Science, University of Illinois at Urbana-Champaign. He received a BS from National Taiwan University and PhD from Stanford University, in Electrical Engineering. His research addresses large scale information access, focusing on “entity-centric” information search, integration and mining over Web and social networks. He received a VLDB Best Papers Selection in 2000, an NSF CAREER Award in 2002, an NCSA Faculty Fellow Award in 2003, IBM Faculty Awards in 2004 and 2005, Academy for Entrepreneurial Leadership Faculty Fellow Award in 2008, and the Incomplete List of Excellent Teachers at University of Illinois in 2001, 2004, 2005, 2006, 2010, and 2011. He loves to bring research results to the real world and, with his students, co-founded Cazoodle, a startup from the University of Illinois, for deepening vertical “data-aware” search over the web.

3. Duration and Sessions

Proposed duration of the tutorial is 3 hours. The entire tutorial will be divided into three sessions: Event Detection, Event Description and User Profiling.

4. Topic and Description

Here is a brief outline of the tutorial.

- Event Detection for Microblogging Platforms
  - Event Detection using Tweet Content: Bursty keywords, Graph community analysis, Locality sensitive hashing, Conditional random fields, Hashtags, Tag correlations, Segments (or semantic phrases) [1, 4, 13, 18]
  - Applications of Event Detection: Forest-fires, Sporting events, Local festivals, Drug related adverse events, Traffic events, Epidemics, Earthquakes, Emerging controversial events [7, 12, 20, 23]
• Event Description for Microblogging Platforms
  – Finding Best Phrase to Summarize an Event [24]
  – Finding Event Types [21]
  – Finding Event Timespans [19]
  – Finding Event Credibility [2, 9]
• User Profiling for Microblogging Platforms
  – Content Based Profiling: General Content Based Attribute Profiling (Gender, Interests), Location Profiling [3, 5, 8, 10, 17]
  – Network Based Profiling: General Network Based Profiling (Interests, University), Location Profiling [6]
  – Hybrid Approaches [14, 16, 22]
  – Co-Profiling Attributes and Relationships [15]

5. Audience

Researchers in the field of analysis of microblogging platforms will benefit the most as this will give them an exhaustive overview of the research in the direction of event detection for microblogging platforms. 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. Folks from the industry will benefit from the discussions on a large number of applications where such mechanisms have already been applied.

After the tutorial, the audience will be able to appreciate and understand the following: (1) How text streams can be processed to identify interesting topics (or events); (2) Challenges in displaying those topics (or events) to users; (3) Challenges in predicting location of events; and (4) Challenges in user profiling for microblogging platform users.

6. Relevance to WWW 2014 Areas

The proposed tutorial is most related to the “Web mining” and the “Social networks and graph analysis” areas of WWW 2014. Besides analysis of other forms of data, both the areas deal with analysis of the Twitter network and data in the form of tweets.

7. References