# DUBMOD'14 - International Workshop on Data-driven User Behavioral Modeling and Mining from Social Media

Jalal Mahmud, Jeffrey Nichols, Michelle Zhou
IBM Research – Almaden
650 Harry Rd, San Jose, CA 95120
{jumahmud, jwnichls, mzhou}@us.ibm.com

James Caverlee
Texas A&M University
College Station, TX
caverlee@cse.tamu.edu

Yi Zeng
Chinese Academy of Sciences
China
yi.zeng@ia.ac.cn

Liang Chen
Zhejiang University
China
cliang@zju.edu.cn

John O'Donovan
University of California, Santa Barbara
Santa Barbara, CA
jod@cs.ucsb.edu

## ABSTRACT
Massive amounts of data are being generated on social media sites, such as Twitter and Facebook. These data can be used to better understand people (e.g., personality traits, perceptions, and preferences) and predict their behavior. As a result, a deeper understanding of users and their behavior can benefit a wide range of intelligent applications, such as advertising, social recommender systems, and personalized knowledge management. These applications will also benefit individual users themselves and optimize their experience across a wide variety of domains, such as retail, healthcare, and education. Since mining and understanding user behavior from social media often requires interdisciplinary effort, including machine learning, text mining, human-computer interaction, and social science, our workshop aims to bring together researchers and practitioners from multiple fields to discuss the creation of deeper models of individual users by mining the content that they publish and the social networking behavior that they exhibit.

## Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications - Data Mining; J.4 [**Computer Application]**: Social and Behavioral Sciences

## General Terms
Algorithms, Design, Experimentation.

## Keywords
User Modeling; Social Media Analysis; Data-driven.

## 1. INTRODUCTION
People from all walks of life are using social media sites, such as Twitter and Facebook, to share data about social events, express opinions, discuss their interests, publicize businesses, recommend products, and, explicitly or implicitly, reveal personal information. Many research projects and commercial products have used this information to drive a variety of services, such as recommending local news stories [2], detecting events [3], determining group sentiment of products and brands [1], and smart crowd-sourcing [8]. The majority of this work, however, has focused on aggregating data across many users. In contrast, our focus is to create deep models of individual users from social media, including the content users publish and the behavior that they exhibit. This has been an active area of research recently and a number of papers on this topic have appeared in related conferences and workshops [4, 5, 6]. Various analytic techniques have been applied to infer users' home location [7], gender [4, 6], age [4], regional origin [4], ethnicity [5], political orientation [4, 5], and personality [9]. Although there is abundant work in this area, there are many technical challenges remaining to be addressed. Our workshop aims to bring together researchers and practitioners who are interested in this topic to share their experiences and discuss remaining challenges. While acquiring a deep understanding of users brings great benefits to businesses and individual users themselves, such an understanding may also reveal information that users would prefer to keep private. Such concerns are particularly important because individuals do not have complete control over the information they share about themselves; friends of a user may inadvertently divulge private information about that user in their own posts.

This workshop builds upon the success of our previously held workshops in conjunction with ACM IUI 2012, ACM CIKM 2012 and ACM CIKM 2013 and focuses more on data-centric analysis and mining aspects of user and user behavior understanding from social media. In particular, we plan to solicit work on topics in one or more of the following areas:

**Analytics**

- Interactive visual analytics for understanding derived user models
- Integrated content and social network mining and analysis to derive user insights
- Mining of heterogeneous data sources to derive user insights
- Discovery of trustworthy information and information sources

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*CIKM*'14, November 3-7, 2014, Shanghai, China.
Copyright 2014 ACM 978-1-4503-2598-1/14/11…$15.00.
http://dx.doi.org/10.1145/2661829.2663538

- Identification of significant user attributes (e.g., personality or preferences) for task-specific user behavior modeling and prediction (e.g., information collection or spreading tasks)
- Predictive analytics of user behavior
- User behavior modeling and mining at scale using Big Data approaches

**Feasibility/challenges of understanding individual users from social media**

- What aspects of an individual can be modeled from their public social media postings?
- What aspects cannot be modeled?
- What aspects should not be modeled?
- How accurate are the models that can be extracted?
- What are the best techniques to model user behavior?
- How might the creation of such models be thwarted? (e.g. to preserve privacy while still allowing participation on a social network)

**Protecting user privacy**

- What information about a user can be modeled while keeping the sensitive information private?
- How can users monitor what information has been revealed about themselves on social media and obfuscate any sensitive information that has been accidently revealed?

**Domains and Applications**

- Domain-specific user modeling using public social media including Twitter, Facebook, MySpace, social Q&A sites, and Amazon.com reviews for
  - Retail
  - Healthcare
  - Education
  - Sports
  - News
- Domain-independent user modeling using public social media such as twitter, Facebook, MySpace, and Foursquare to derive a wide variety of user traits including:
  - Locations
  - Behaviors/Personality
  - Demographics
  - Age
  - Gender
- Enterprise-focused user modeling using social media data on public social networks and communications (e.g., emails and blogs) within an enterprise
  - Employees' social and collaboration patterns in a workplace
  - Work-related personality traits such as innovativeness, flexibility, and adaptiveness
- Task-specific user modeling for
  - Information recommendation
  - Crowd-sourcing
  - Expert finding
  - Social Q&A

## 2. OBJECTIVE

This workshop aims to achieve three main objectives:

- Identify key research issues and challenges in designing and developing techniques for modeling and mining users and their behavior from social media; and
- Establish and grow a community that consists of researchers and practitioners from multiple disciplines to tackle the difficult problems of user behavior understanding from social media; and
- Initiate collaboration (e.g., creating and sharing public datasets) among different teams.

We hope to bring together researchers and practitioners from diverse areas, such as user modeling, social media analysis, natural language processing, data mining, machine learning, privacy and security, to discuss these issues and share results.

## 3. FORMAT

We will hold a half-day workshop program on Nov, 3, 2014. The program begins with a welcome note. After that there will be a madness session during which participants introduce themselves and their work in 5 minutes. Participants will be encouraged to make slides for the madness session, but this will not be required. Each paper session will conclude with a discussion led by a pre-chosen workshop participant. These discussions will tie together common themes of the presentations and hopefully lead to insightful discussions about further research directions. The program will end with a open discussion where participants will discuss the current state of the art, focus areas, and opportunities for future research.

## 4. ORGANIZERS


- Jalal Mahmud, IBM Research – Almaden, USA, jumahmud@us.ibm.com
- James Caverlee, Texas A&M University, USA, caverlee@cse.tamu.edu
- Jeffrey Nichols, IBM Research – Almaden, USA, jwnichls@us.ibm.com
- John O'Donovan, University of California, Santa Barbara, USA, jod@cs.ucsb.edu
- Michelle Zhou, IBM Research - Almaden, USA, mzhou@us.ibm.com
- Yi Zeng, Chinese Academy of Sciences, China. yi.zeng@ia.ac.cn
- Liang Chen, Zhejiang University, China, cliang@zju.edu.cn


## 5. REFERENCES


[1] http://twittersentiment.appspot.com

[2] Owen Phelan, Kevin McCarthy, and Barry Smyth. (2009). Using Twitter to Recommend Real-time Topical News. RecSys.

[3] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In WWW, 2010.

[4] Delip Rao, David Yarowsky, Abhishek Shreevats, Manaswi Gupta, Classifying Latent User Attributes in Twitter, In Proc. of SMUC '10.

[5] Marco Pennacchiotti and Ana-Maria Popescu, A Machine Learning Approach to Twitter User Classification, In Proc. of ICWSM'11.

[6] Popescu, A. and Grefenstette, G. Mining User Home Location and Gender from Flickr Tags. ICWSM '10.

[7] Brent Hecht, Lichan Hong, Bongwon Suh, Ed H. Chi, Tweets from Justin Bieber's Heart: The Dynamics of the "Location" Field in User Profiles, CHI 2011.

[8] Jeffrey Nichols and Jeon-Hyung Kang, Asking questions of targeted strangers on social networks, In Proc. of CSCW 2012.

[9] Thin Nguyen, Dinh Phung, Brett Adams, Svetha Venkatesh, Towards Discovery of Influence and personality traits through social link prediction, In Proc. of ICWSM 2011