

Modeling Temporal Dynamics and Geographical Language Variations of Social Streams

Liangjie Hong
Lehigh University

11/12/2012 at 11:00 am
CBIM multipurpose room (Room 22)

Abstract

In recent year, social media services have become indispensable communication tools for online users for disseminating breaking news, eyewitness accounts, individual expression, and protest groups. On one hand, these services are considered as important complementary media sources to traditional media outlets, such as televisions, newspapers and magazines. Indeed, the dynamics between these media sources attract researchers to reveal interesting patterns on how news stories and topics evolve over time across multiple sources. On the other hand, rich meta-data augmented social media also provides unique opportunities to understand user behaviors that cannot be easily studied in conventional settings. For instance, some social media services allow users to associate generated content with geographical locations. This information opens a set of interesting questions, such as whether we can predict locations of users based on the generated content.

In this talk, I focus on tackling the challenges mentioned above, discussing how to model temporal dynamics of topics and geographical language variations of social streams. These two research directions represent some exceptional characteristics of social streams which is useful for many applications in recommendation and personalization. For temporal modeling, we address the problem of modeling text streams from two news sources - Twitter and Yahoo! News, extending standard topic models by allowing each text stream to have both local topics and shared topics. We associate each topic with a time-dependent function that characterizes its popularity over time. By integrating the two models, we effectively model the temporal dynamics of multiple correlated text streams in a unified framework, discovering interesting transfers of topics over two data sources and capturing the rise and fall of topics over time. For geographical modeling, we present an algorithm by modeling diversity in Twitter messages based on topical diversity, geographical diversity, and an interest distribution of the user. Our model exploits

sparse factorial coding of the attributes, thus allowing us to deal with a large and diverse set of covariates efficiently. The proposed approach is vital for applications such as user profiling, content recommendation and topic tracking. We show high accuracy in location estimation based on our model. Moreover, the algorithm identifies interesting topics based on location and language.

Bio

Liangjie Hong is a Ph.D. candidate in the Dept. of Computer Science and Engineering at Lehigh University. His research interest is applied machine learning for understanding user behaviors in social media, exploring and modeling rich user-generated content, with applications in recommendation and personalization. He has interned twice in Yahoo! Labs in 2010 and 2011, and once in LinkedIn Corp. in 2011. During the time period of the Ph.D. study, he has published more than 10 technical papers, including 2 in SIGIR, 3 in KDD, 1 in WWW, 1 in WSDM and 1 in AAAI. He received his M.S. in 2010 from Lehigh University and B.S. in 2007 from Beijing University of Chemical Technology.

Faculty Host: Tina Eliassi-Rad