NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **Learning and Modeling the Underlying Semantics of Online Discussions**

Speaker: Mr Ghasem Heyrani Nobari

Date/Time: 17 September 2014, Wednesday, 02:00 PM to 03:30 PM

Venue: MR1, COM1-03-19

Supervisor :   Dr Chua Tat Seng, Professor, School of Computing

Abstract:

User's online activities and their participation in social networks have increased significantly in recent years. As a result of this massive participation, the amount of user-generated contents (UGC) has also been expanded rapidly. There are many types of UGC; they can be as short as a comment, tweet or a Facebook post or as long as a review or blog post. In line with this rapid development of social networks, online discussions are growing as a popular and effective source of information for users because of their timely, lively and flexible content. In recent years, this popularity and user demands have resulted in substantial growth in both the number of users and diversities of topics among online discussions. Usually online discussions are developed and advanced incrementally by groups of users with various backgrounds and intents. However the flexibility, fast updating, informal languages and dynamic structure of online discussions make them a challenge for new users and automated systems to understand and learn the underlying semantics of these long threaded discussions.

Recently, many approaches have been proposed to model relations of topics and learn the semantics from user-generated contents. These approaches are usually based on Bayesian models, clustering methods and different types of topic models. However, the problem of learning and modeling online discussions is a complex and multi-faceted problem which requires a mixture of these methods to analyze and model the complex and dynamic relations between discussions, topics and users. Therefore, for accurate and comprehensive modeling of online discussions, we need a unified framework that can learn the hidden relations between all these three aspects of users, topics and discussions. To learn all these relations and hidden structures, in this thesis we proposed a unified framework as follows: (a) We proposed a novel unsupervised Aspect-Action topic model (AS-AC), that enable us to identify primary topics and their dependencies from a sequence of user posts. In particular, we jointly modeled aspects with their associated actions to boost the precision of our generative process, where actions play the role of defining the functionalities for a group of aspects. (b) We extended the AS-AC model to learn and generate the underlying

hierarchical structure of a discussion. We utilized a fast binary approach that were able to capture the evolution of topics and subtopics accurately in the duration of a discussion. (c) We presented a model for automatic identification of user's objectives and intents within a discussion which can be used for extensive evaluation and comparison between discussions. This will enable us to calculate the semantic similarity between any two sub-topics by generating the aspect-action relationship graph and finding the group of highly connected topics. In this thesis, we conduct experiments on Apple discussion forums based on various products with over 3.3 million user posts from 300k discussion threads. Our evaluation indicates that the joint aspect-action model results in substantial improvements in accuracy of discussion modeling and capturing latent relations between users, posts and topics.