香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

for favour of posting

PhD Candidate:	Mr Kai LIU
Supervisor:	Dr Kwok Wai CHEUNG
External Examiner:	Prof Carson Kai Sang LEUNG Dr Raymond Hau San WONG
Time:	19 Aug 2014 (Tuesday) 10:00 am – 12:00 nn (35 mins presentation and 15 mins Q & A)
Venue:	RRS732, Sir Run Run Shaw Building, HSH Campus

"Detecting Stochastic Motifs in Network and Sequence Data for Human Behavior Analysis"

Abstract

With the recent advent of web 2.0, mobile computing, and pervasive sensing technologies, human activities can readily be logged, leaving digital traces of different forms. For instance, human communication activities recorded in online social networks allow user interactions to be represented "network" data. Also, human daily activities can be tracked in a smart house, where the log of sensor triggering events can be represented as "sequence" data. This thesis research aims to develop computational data mining algorithms using the generative modeling approach to extract salient patterns (motifs) embedded in such network and sequence data, and to apply them for human behavior analysis.

Motifs are defined as the recurrent over-represented patterns embedded in the data, and have been known to be effective for characterizing complex networks. Many motif extraction methods found in the literature assume that a motif is either present or absent. In real practice, such salient patterns can appear partially due to their stochastic nature and/or the presence of noise and thus a probabilistic approach is used to model motifs.

For network data, we use a probability matrix to represent a network motif and propose a mixture model to extract network motifs. A component-wise EM algorithm is adopted where the optimal number of stochastic motifs is automatically determined with the help of a minimum message length criterion. Considering also the edge occurrence ordering within a motif, we model a motif as a mixture of first-order Markov chains for the extraction. Using a probabilistic approach similar to the one for network motif, an optimal set of stochastic temporal network motifs are extracted. We carried out rigorous experiments for evaluating the performance of the proposed motif extraction algorithms using both synthetic data sets and real-world social network data sets and mobile phone usage data sets, and obtained promising results. Also, we found that some of the results can be interpreted using the social balance and social status theories which are well-known in social network analysis.

To study if the use of stochastic temporal network motifs can allow human behaviors to be not only better characterized but also predicted, we make use of stochastic temporal network motifs as some local structure features of the network and incorporate them into a factor graph model for followee recommendation prediction (essentially a regression problem) in online social networks. The factor graph model with the stochastic temporal network motifs incorporated is found to outperform significantly the existing state-of-the-art methods for the prediction task.

For sequence data, the probabilistic framework proposed for the stochastic temporal network motif extraction can be adopted. One possibility is to make use of the edit distance in the probabilistic framework so that the subsequences with minor ordering variations can first be grouped as the initial motif candidates so that the optimal set of temporal motifs can be determined. We applied this approach to extract temporal motifs from a smart home data set which contains a log of sensor triggering events corresponding to some activities performed by residents in the smart home. The unique behavior extracted for each resident based on the detected motifs is also discussed.

*** ALL INTERESTED ARE WELCOME ***