

PHD DISSERTATION DEFENSE ANNOUNCEMENT

Structural Exploration and Inference of the Network

Ruihua Cheng

Ph.D. Candidate in Probability and Applied Statistics

Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ

Advisor: Dr. Ji Meng Loh, Associate Professor, Department of Mathematical Sciences, NJIT

Co-Advisor: Dr. Zhi Wei, Associate Professor, Department of Computer Science, NJIT

Committee Members:

Dr. Antai Wang, Associate Professor, Department of Mathematical Sciences, NJIT

Dr. Yixin Fang, Assistant Professor, Department of Population Health, NYU

Dr. Yi Chen, Associate Professor, School of Management, NJIT

Date: Thursday, July 28, 2016 Time: 11:00 AM - 1:00 PM Venue: Cullimore Hall, Room 611

ABSTRACT

This thesis consists of two parts. In the first part, a learning-based method for classification of online reviews that achieves better classification accuracy is extended. Automatic sentiment classification is becoming a popular and effective way to help online users or companies to process and make sense of customer reviews. The method combines two recent developments. First, valence shifters and individual opinion words are combined as bigrams to use in an ordinal margin classifier. Second, relational information between unigrams expressed in the form of a graph is used to constrain the parameters of the classifier. By combining these two components, it is possible to extract more of the unstructured information present in the data than previous methods, like support vector machine, random forest, hence gaining the potential of better performance. Indeed the results show a higher classification accuracy on empirical real data with ground truth as well as on simulated data.

The second part deals with graphical models. Gaussian graphical models are useful to explore conditional dependence relationships between random variables through estimation of the inverse covariance matrix of a multivariate normal distribution. An estimator for such models appropriate for multiple graphs analysis in two groups is developed. Under this setting, inferring networks separately ignores the common structure, while inferring networks identically would mask the disparity. A generalized method which estimates multiple partial correlation matrices through linear regressions is proposed. The method pursues the sparsity for each matrix, similarities for matrices within each group, and the disparities for matrices between groups. This is achieved by a l_1 penalty and a l_2 penalty for the pursuit of sparseness and clustering, and a metric that learns the true heterogeneity through optimization procedure. Theoretically, the asymptotic consistency for both constrained l_0 method and the proposed method to reconstruct the structures is shown. Its superior performance is illustrated via a number of simulated networks. An application to polychromatic flow cytometry data sets for network inference under different sets of conditions is also included.