Colloquium

107年09月26日(三) 14:30 應用數學系多媒體教室（理408室）

On the Price of Anonymity in Heterogeneous Statistical Inference

Abstract:

Statistical inference is a fundamental task in data science, where a decision maker aims to determine a hidden parameter based on the data it collects, as well as how the probability distribution of the data depends on the target parameter. In many modern applications such as crowdsourcing and sensor networks, data is heterogeneous and collected from various sources, each of which follows different distributions. These sources, however, may be anonymous to the decision maker due to considerations in identification costs and privacy. Since the distribution becomes unknown, it is unclear what is the impact of such anonymity on the performance of statistical inference, and how to carry out optimal inference. In this talk, I will present our recent work towards settling this question, focused on binary hypothesis testing. Considering the anonymity of data sources, it is natural to formulate it as a composite hypothesis testing problem. First, we propose an optimal test called *mixture likelihood ratio test*, a randomized threshold test based on the ratio of the uniform mixture of all the possible distributions under one hypothesis to that under the other hypothesis. Second, we focus on the Neyman-Pearson setting and characterize the error exponent of the worst-case type-II error probability as the dimension of data tends to infinity while the proportion among the dimensions of different data sources remains constant. It turns out that the optimal exponent is a generalized divergence between the two families of distributions under the two hypotheses. Our results elucidate the price of anonymity in heterogeneous hypothesis testing and can be extended to more general inference tasks.

This talk is based on the joint work with my student Wei-Ning Chen.

王奕翔 教授

Professor I-Hsiang Wang

國立台灣大學電機工程學系
助理教授

Ph.D.: Electrical Engineering and Computer Sciences, University of California, Berkeley, 2011

Research areas are:
1. information theory
2. networked information and data processing

Bio:
I-Hsiang Wang received his Ph.D. in Electrical Engineering and Computer Sciences from University of California at Berkeley, USA, in 2011. From 2011 to 2013, he was a postdoctoral research associate in the School of Computer and Communication Sciences (IC) at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. In Fall 2013, he joined National Taiwan University, where he is now an assistant professor. Prof. Wang's expertise lies in information theory, statistical learning, and networked information and data processing. He received the Berkeley Vodafone Fellowship in 2006 and 2007. He was a finalist of the Best Student Paper Award of IEEE International Symposium on Information Theory, 2011. He won the 2017 IEEE Information Theory Society Taipei Chapter and IEEE Communications Society Taipei/Tainan Chapters Best Paper Award for Young Scholars, and the 2016 National Taiwan University Distinguished Teaching Award (top 1%). He served on the technical program committees of flagship conferences in information theory, including IEEE International Symposium on Information Theory (ISIT) and IEEE Information Theory Workshop (ITW).

國立高雄大學
應用數學系

www.math.nuk.edu.tw