Title: **Information theoretic learning in pattern classification (3 hours)**

Speaker: Baogang Hu

Affiliation: National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China

E-mail: hubg@nlpr.ia.ac.cn

Abstract:

In this tutorial, I will start with my personal view on the study of machine learning. The issues can be considered as four modules connected in a hierarchical way as “what to learn?”, “how to learn?”, “what to evaluate?” and “what to adjust?”. The first issue, also called “learning target selection”, does not receive sufficient recognitions within our community if compared with the existing investigations on the subject of “feature selection”. The tutorial will present the “information theoretical learning” (also termed ITL by Principe, et al. 2000, 2010) in relation to the issues. The objective of the tutorial is to demonstrate that ITL will not only present a fundamental understanding to the learning target selection, but also lead to new classification tools in machine learning.

The tutorial will focus on pattern classification in the basis on ITL. I will introduce the novel theory of abstaining learning for both Bayesian classifiers and mutual information classifiers. Abstaining, or a reject option in classification, is one of the most important behaviors in real-life decision making from humans, which may significantly reduce total cost or risk in applications. Based on the theory, I will introduce the cost-free learning from the real-world data sets in comparison with cost-sensitive learning. The significance of the cost-free learning is demonstrated in the background of class-imbalance problems when costs are unknown for both errors and rejects. The connections between the empirical measures and information measures are presented. The fundamental relations are upper bound and lower bound for both Bayesian error and non-Bayesian error with respect to conditional entropy in binary classifications. I will also demonstrate 24 information measures in the evaluation of abstaining binary classifications. The tutorial is concluded by the further discussions on the emergences of abstaining learning and cost-free learning in the context of “big-data” classifications.

Related Publications:

1. Hu, B.-G. and Wang, Y., “Evaluation Criteria Based on Mutual Information for Classifications Including Rejected Class,” Acta Automatica Sinica, vol. 34, pp. 1396-1403, 2008.
2. Xing, H.-J. and Hu, B.-G., “Two-Phase Construction of Multilayer Perceptrons Using Information Theory”, IEEE Transactions on Neural Networks, Vol. 20, pp. 715-721, 2009.
3. Yuan, X.-T. and Hu, B.-G., “Robust Feature Extraction via Information Theoretic Learning”, The 26th International Conference on Machine Learning (ICML 2009).
4. He, R., Zheng, W.-S. and Hu, B.-G., “Maximum Correntropy Criterion for Robust Face Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 1561-1576, 2011.
5. He, R., Zheng, W.-S. and Hu, B.-G., X.-W. Kong, “A Regularized Correntropy Framework for Robust Pattern Recognition”, Neural Computation, Vol. 23, pp. 2074-2100, 2011.
6. He, R., Hu, B.-G., Zheng, W.-S. and Kong, X.-W., “Robust Principal Component Analysis based on maximum Correntropy criterion”, IEEE Transactions on Image Processing, Vol. 20, pp. 1840-1494, 2011.
7. Hu, B.-G., He, R., and Yuan, X.-T.,“Information-theoretic measures for objective evaluation of classifications,” Acta Automatica Sinica, vol. 38, pp. 1160–1173, 2012.
8. Hu, B.-G., “What are the Differences between Bayesian Classifiers and Mutual-Information Classifiers?”, IEEE Transactions on Neural Networks and Learning Systems. Vol. 25, pp. 249-264, 2014.
9. Zhang, X., and Hu, B.-G., “A New Strategy of Cost-Free Learning in the Class Imbalance Problem”, IEEE Transactions on Knowledge and Data Engineering，Vol. 26, pp. 2872-2885, 2014.
10. He, R., Hu, B.-G., Yuan, X.-T., and Wang, L., Robust Recognition via Information Theoretic Learning, Springer, 2014.
11. Hu, B.-G., “Information Theory and its Relation to Machine Learning”, January 18, 2015. http://arxiv.org/abs/1501.04309).
12. Hu B.-G., and H.-J. Xing, “An Optimization Approach of Deriving Bounds between Entropy and Error from Joint Distribution: Case Study for Binary Classifications”, Entropy, Vol. 18, pp. 1-19, 2016.

Bio:

Dr. Bao-Gang Hu is currently a full Professor with NLPR (National Laboratory of Pattern Recognition), Institute of Automation, Chinese Academy of Sciences, Beijing, China. He received his M.S. degree from the University of Science and Technology, Beijing, China in 1983, and his Ph.D. degree from McMaster University, Canada in 1993. From 2000 to 2005, he was the Chinese Director of LIAMA (the Chinese-French Joint Laboratory supported by CAS and INRIA). His current research interests are pattern recognition and computer modeling.