

Feature and Sample Reduction for Classification Problems

Abstract

A classification system such as a neural network maps input data characterized by a number of features onto output classes. Successful deletion of “irrelevant or unimportant” features and samples in the training set, without sacrificing the classification accuracy, could reduce network complexity and learning effort. Such a reduction technique is highly desirable for many application problems. In this talk a comparison on a number of well-known techniques based on the principal component analysis, the mutual information, the support vector machines and the neural network sensitivity analysis will be presented. We shall present a proposal to develop a feature and sample selection method for supervised multi-classification problems using Sensitivity Measures for an ensemble of multilayer feedforward neural networks (Multilayer Perceptrons or Radial Basis Function Neural Networks). This proposed technique is based on some recent results on generalization error locally near the training points. A number of experimental results using datasets such as the UCI, the 99 KDD Cup, and the text categorization, will be presented.

About the Speaker

Daniel S. Yeung received the Ph.D. degree in applied mathematics from Case Western Reserve University. In the past, he has worked as an Assistant Professor of Mathematics and Computer Science at Rochester Institute of Technology, as a Research Scientist in the General Electric Corporate Research Center, and as a System Integration Engineer at TRW, all in the United States. He was the chairman of the department of Computing, The Hong Kong Polytechnic University, Hong Kong where now he is a Chair Professor. His current research interests include neural-network sensitivity analysis, data mining, Chinese computing, and fuzzy systems. He was the Chairman of IEEE Hong Kong Computer Chapter (91 and 92), an associate editor for both IEEE Transactions on Neural Networks and IEEE Transactions on SMC (Part B). He is a member of the Board of Governor of the IEEE SMC Society, and the Vice President for Technical Activities for the same Society. He co-founded and served as a General Co-Chair since 2002 for the International Conference on Machine Learning and Cybernetics held annually in China. He leads a group of researchers in Hong Kong and China who are actively engaging in research works on computational intelligence and data mining. He is also the founding Chairman of the IEEE SMC Hong Kong Chapter. Professor Yeung is a Fellow of the IEEE.