

# **Diversified Sensitivity based Undersampling for Imbalance Classification Problems**

**Daniel Yeung**

## **Abstract**

Undersampling is a widely adopted method to deal with imbalance pattern classification problems. The problem of class imbalance occurs in many pattern classification problems like network security, medical diagnostics and digit recognition. Current methods mainly depend on either random resampling on the majority class or resampling at the decision boundary. Random based undersampling fails to take into consideration informative samples in the data while resampling at the decision boundary is sensitive to class overlapping. Both techniques ignore the distribution information of the training dataset. In this talk a diversified sensitivity based undersampling method is proposed to address this problem. Samples of the majority class are clustered to capture the distribution information and enhance the diversity of the resampling. A stochastic sensitivity measure is applied to select samples from both clusters of the majority class and the minority class. By iteratively clustering and sampling, a balanced set of samples yielding high classifier sensitivity is selected. We will present experimental results on training sample selection by various methods, and analysis of their performances on both artificial and UCI datasets with both imbalance situations and different ratios of overlapping.