香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Ms. LI Shuxian
Date	23 August 2024 (Friday)
Time:	10:00 am – 12:00 noon (35 mins presentation and 15 mins Q & A)
Venue:	ZOOM (Meeting ID: 910 1065 1265) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 12:00 nn, 22 August 2024)

Multi-class Imbalance Learning: Beyond Class Imbalance Ratios

Abstract

Class imbalance issues often arise in real-world scenarios, especially in fields like video categorization, human behavior analysis, and medical decision-making. Class imbalance, where some classes have fewer training samples than others, can lead to lower classification performance, especially on minority classes. The challenge becomes even more complicated when there are multiple classes with imbalanced distributions.

Current techniques for tackling multi-class imbalance problems include the decomposition approach, data sampling approach, cost-sensitive approach, and ensemble methods. These techniques always focus on the information of class imbalance ratios when dealing with multi-class imbalance problems without considering other useful data properties to improve the classification model.

With this in mind, this thesis offers three novel approaches to deal with the multi-class imbalance issue more efficiently. First, we present AdaBoost.AD, an ensemble model with adaptive sample weights determined based on different data properties. The first data property is the class imbalance ratios between classes. Beyond class imbalance ratios, we also consider the data density and the classification difficulty, both of which are encoded into the ensemble model. Second, we create ARConvL, a convolutional learning framework with adaptive loss functions. This framework can trace the changing latent feature space, learning the feature distribution throughout training epochs. Then, adaptive loss functions are determined based on the feature distribution, which can optimize the distribution of each class under the class imbalance scenario and enlarge the discrimination between different classes. Finally, we propose BEDCOE, an online ensemble model with borderline enhanced strategy and disjunct cluster based oversampling method. Beyond class imbalance ratios, this proposed approach first considers the borderline data property, based on which to enhance the training times for training samples with lower imbalance ratios and closer to the borderline area. Then, based on the disjunct data property, synthetic samples are generated to deal with the online multi-class imbalance issue.

In each case, we provide not only theoretical support but also empirical evidence of the effectiveness of these methods. We show that our proposed techniques can significantly improve the performance of multi-class imbalance learning.

***** ALL INTERESTED ARE WELCOME *****