Imbalanced Classification: Common Approaches and Open Problems

Francisco Herrera
University of Granada, Spain
herrera@decsai.ugr.es

Short Abstract

Classifier learning with data-sets which suffer from imbalanced class distributions is an important problem in data mining. This issue occurs when the number of examples representing one class is much lower than the ones of the other classes. Its presence in many real-world applications has brought along a growth of attention from researchers.

The aim of this talks is to shortly review the two common approaches to dealing with imbalance, sampling and cost sensitive, and the use of ensemble techniques based on the mentioned approaches.
We will pay special attention to some open problems, in particular we will discuss the multi-class imbalance problems, the overlapping between the classes and the data fracture between training and test distribution that provokes the bad behavior of classifiers.