

# Efficient Fused Learning for Distributed Imbalanced Data

Jie Zhou

*Capital Normal University*  
*E-mail: zhoujie@amss.ac.cn*

**Abstract:** Any data set exhibiting an unequal or highly-skewed distribution between its classes/categories can be regarded as imbalanced data.

Due to privacy concern and other technical limitations, imbalanced data distributed across locations/machines cannot be simply combined and stored in a single central location. The common used naive averaging estimate may be unstable for imbalanced data. In this paper, we propose a fused estimation for logistic regression in analyzing distributed imbalanced data by combining all the cases available on all machines, which is stable and efficient. The consistency and asymptotic normality of the proposed estimator are established under regularity conditions. Asymptotic efficiency compared with the oracle estimator based on the entire imbalanced data is also studied. Extensive simulation studies show that the proposed estimator is as efficient as the oracle estimator in various situations.