Balancing Misclassification Costs (BMC) in Imbalanced Classification

Fu, Sophia (School: Carmel High School)

Classification tasks in machine learning, essential for applications ranging from fraud detection to medical diagnoses, frequently encounter the challenge of imbalanced datasets. These imbalances can skew predictions towards the majority class, risking oversight of vital minority instances and carrying significant real-world consequences. Established methods, such as Logistic Regression, Support Vector Machines, and ensemble techniques, offer solutions to classification challenges but often struggle with imbalanced datasets. Conventional strategies like resampling and cost-sensitive learning provide value but come with issues like overfitting, data loss, and increased computational demands. A notable disconnect also exists between estimation procedures and evaluation metrics, further complicating the task of accurately gauging model performance. In this work, I present the Balancing Misclassification Costs (BMC) algorithm, an innovative approach designed to adeptly tackle the challenges posed by imbalanced datasets. My method integrates misclassification costs within a unified optimization framework. Capitalizing on rigorous theoretical proof, I have also devised an efficient estimation procedure. Through detailed simulations and its application to a cancer diagnostic dataset, I underscore BMC's superiority over conventional methodologies.

Awards Won:
Second Award of $2,000
China Association for Science and Technology (CAST): Award of $1,200