Topology-Guided Clustering and Local Deep Models for Imbalanced Data Classification

Loai AbdAllah

The Department of Information Systems, The Max Stern Yezreel Valley College, Israel

Abstract:

Imbalanced classification is a persistent challenge in machine learning, where conventional models often fail to adequately represent the minority class due to biased decision boundaries learned from skewed data distributions. While existing techniques such as resampling, cost-sensitive learning, and ensemble methods attempt to mitigate this issue, they typically operate on a global scale, overlooking the underlying structure and local complexity of the feature space.

This paper introduces **TG-LEIC** (*Topology-Guided Local Expert Imbalanced Classifier*), a general and modular framework that reformulates the imbalanced learning problem from a local topology-aware perspective. The core idea is to partition the data space into clusters that reflect its intrinsic structure and train specialized classifiers on rebalanced subsets within each cluster. During inference, each test instance is routed to the most relevant local model based on its proximity to cluster centroids or latent representations.

The framework is algorithm-agnostic and supports interchangeable components, including feature extractors, clustering algorithms, balancing strategies, and classifiers. By aligning the training process with localized data characteristics, TG-LEIC significantly improves minority class recognition and overall classification robustness.

Extensive experiments on benchmark datasets demonstrate that TG-LEIC outperforms traditional global models and popular ensemble-based methods, particularly in terms of minority class recall and F1-score. These results confirm the effectiveness of locally adaptive learning as a principled approach to tackling class imbalance in complex, real-world datasets.