**Efficient Learning on Imbalanced Image set**

**Abstract**

Categorization is a broad range research field that involves many decision-theoretic approaches for identifying data. A large error rate of a classifier is usually associated with the inherent intricacy of the categorization task. However, when the sample size is finite, other aspects, such as small sample size, large number of features, and the intricacy of the categorization rule may also deteriorate classifier performance. This work focuses on handling imbalanced image set.

Tackling imbalanced image sets is a challenging issue being faced by the conventional categorizer. Imbalance problem occur with real world data due to many reasons such as complexity of data, limited availability of data on one class to which the conventional classifiers gets influenced towards major class data. If the data is imbalanced, i.e., data of one class is well represented than the other class, a further demur in performance metrics is expected. For example, in a binary class data, if 99 percent of the data belong to one of two classes, the traditional learning algorithm will still misclassify the remaining 1 percent data that mostly belong to minor class. The proposed approach aim to balance bi-class image set by creating synthetic samples of minority class images.

To evaluate the performance of proposed approach, images from imagenet.com and pexels.com are considered. KNN classification is performed on the image set and evaluated on various evaluation measures. To balance the distribution of images amongst the bi-classed image set, the minority class needs more images. For this, five synthetic image generation techniques are applied on minor class images. Each generation technique produces images that are slightly different in terms of contrast, sketch, brightness etc. The system is retrained with original and synthetic images to evaluate the performance. The results indicate 3% to 4% improvement in the performance and achieves better results against imbalanced image set.

Guided by: Lincy Meera Mathews