Adversarial OverSampling for imbalanced image data classification

Adam Wojciechowski, Mateusz Lango

Motivation
Multi-class imbalance is present in many real world machine learning classification problems. Unfortunately, popular methods for imbalanced data do not bring satisfying effects when applied to image datasets, therefore community concerning computer vision remains in need of efficient methods dealing with low model generalization levels and low multi-class accuracy metrics of underrepresented classes. We present the Adversarial OverSampling (AOS), an algorithm, which simultaneously upsamples and performs dataset augmentation by generating adversarial examples from minority classes during training.

Multi-class imbalance
Multi-class imbalance is a common data characteristic for image classification problems, since it rarely happens, that the number of instances in each class is equal to each other. Researchers focus on two types of multi-class imbalance. Step imbalance is defined by parameters $\mu = \{1, \ldots, N\}$, $C_i$, where:
- $C_i$ is a set of examples in class i
- $N$ is the total number of classes

and $p = \max_i |C_i|$. We call the second type of imbalance the gradual imbalance. It is defined by $p$ and the number of instances in individual classes, that can be approximated by some function. For simplicity, below, we consider linear function, which creates linear imbalance.

Related methods
A Random OverSampling is a method, that can be performed by randomly choosing images from minority classes and copying them into the dataset. There exist more advanced oversampling methods like Synthetic Minority Oversampling Technique (SMOTE), which generates new images by per pixel linear interpolations of original images from the minority classes.

All of the known methods are imperfect when applied to image datasets. ROS is a valid and safe to use method, but it is also primitive, and when used alone, does not bring satisfying improvements. SMOTE can even lead to the degradation of classification performance, which can be partially explained by the fact, that the Euclidean distance used by it is not appropriate for image data.

Adversarial examples
Adversarial example is a general term for a transformed image, in a way, such that it is indistinguishable from original image to the human eye, yet attempt of classifying it results in misclassification. They are used to mislead CNNs often with malicious intentions.

Adversarial OverSampling algorithm
Adversarial OverSampling is a hybrid method for addressing data imbalance, operating on data level as well as algorithmic level. The method needs two parameters: adversarial intensity $\epsilon$ and augmentation intensity $\rho$. The method applies Random OverSampling during pre-training phase, creating base for further transformations. Subsequently, at the beginning of each batch, pick images from minority classes with probability $P_i \sim \mathcal{B}(1, p_i)$ and compute their signed gradients and add them to original images with intensity $\epsilon$. Finalize by forwarding, backpropagating and updating $\Theta$. Repeat.

Results
In our research, we focused on comparing AOS to baseline data BASE and Randomly OverSampled data. Below we report the results of our experiments where the classification performance was measured by F1-score, which consolidates both precision and recall of the CNN, obtained from test sets. In Table 1 we present results on Intel’s ‘Image Scene Classification of Multiclass’ dataset [3] and CIFAR-10, with imbalance parameters, which can be encountered in naturally imbalanced datasets. Let $\rho = 20$ and $\mu = 60$ respectively, $\epsilon = 0.007$ for both.

Table: 1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\mu$</th>
<th>BASE</th>
<th>ROS</th>
<th>AOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-10</td>
<td>0.4</td>
<td>72.68</td>
<td>71.56</td>
<td>73.69</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>0.6</td>
<td>61.32</td>
<td>62.48</td>
<td>66.33</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>0.8</td>
<td>61.12</td>
<td>64.56</td>
<td>65.16</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>1.0</td>
<td>61.05</td>
<td>49.84</td>
<td>51.18</td>
</tr>
</tbody>
</table>

In Table 2 we present more extreme setup to emphasize the differences between performances of those approaches. Let $\rho = 500$ and $\epsilon = 0.1$.

Table: 2

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\mu$</th>
<th>BASE</th>
<th>ROS</th>
<th>AOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINST</td>
<td>0.3</td>
<td>83.39</td>
<td>85.00</td>
<td>92.92</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>0.3</td>
<td>44.66</td>
<td>43.84</td>
<td>45.41</td>
</tr>
<tr>
<td>MINST</td>
<td>0.9</td>
<td>76.38</td>
<td>79.69</td>
<td>83.51</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>0.9</td>
<td>72.68</td>
<td>73.69</td>
<td>75.16</td>
</tr>
</tbody>
</table>

Conclusion
We propose a method for addressing multi-class imbalance applicable to image data and CNNs. Our method proved to bring results superior to compared ones. Therefore, that the imbalance gets, the more efficiently our method handles the imbalance, for given problem.

Bibliography
[3] Intel’s ‘Image Scene Classification of Multiclass’ dataset

In the figure above there is depicted an exemplary set of instances of minority class within one batch. ‘AOS 1’ has $\epsilon = 0.05$ and ‘AOS 2’ has $\epsilon = 0.15$. 

![Image of adversarial examples](image-url)