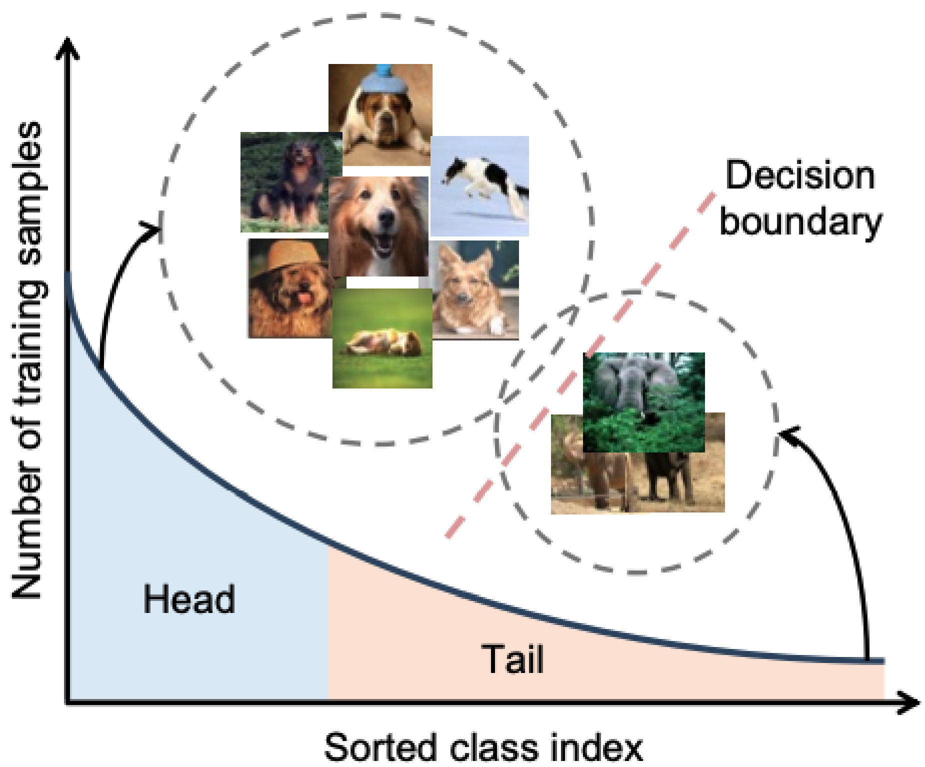
Samer Baslan

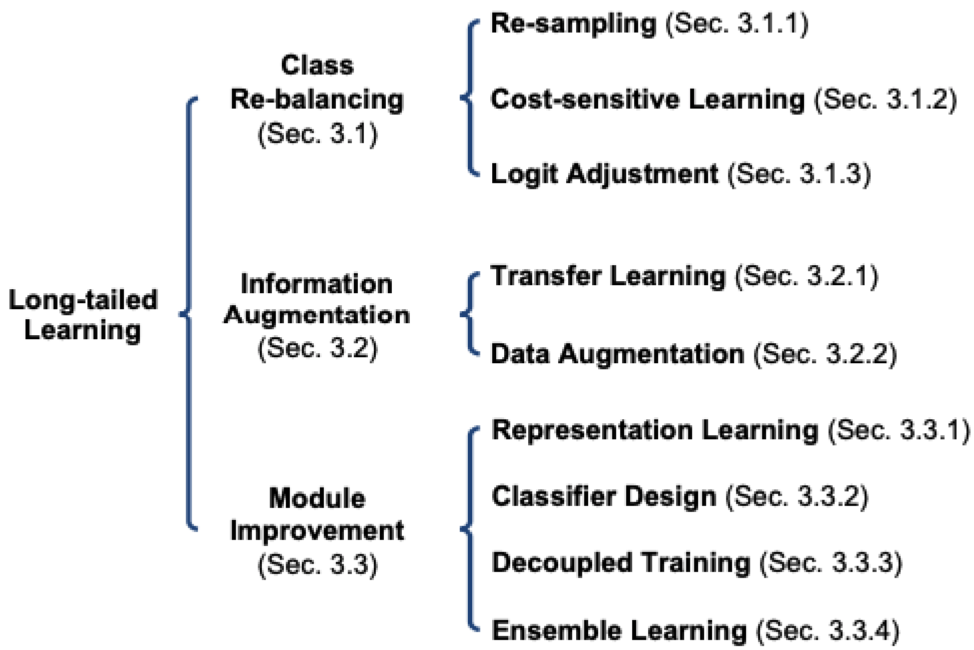
An Introduction to Deep Long-Tailed Learning

Samer Baslan 2 hours ago · 5 min read

This survey by Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan and Jiashi Feng covers the following topic in far grater detail and I highly recommend checking it out for a more thorough discussion the ideas discussed in this article.



however, is that in real world applications, training samples typically a long-tailed class distribution, where a small portion of classes have massive sample points but the others are associated with only a few samples. Thus, a model can be easily biased towards the head classes, resulting in a poor performance on the tail classes [1].



Many methods to counter such class imbalances in the data have been studied, mostly grouped within 3 categories. In this article, I hope to provide a brief summary of these methods.

Class Re-balancing

Class re-balancing is a mainstream paradigm in long-tailed learning that seeks to balance the training sample numbers of different classes during model training [1].

1. **Re-sampling:** Re-sampling is one of the most widely used methods over the last decade for tackling class imbalance issues. The most common types of sampling done are random over-sampling (ROS) and random under-sampling (RUS). The former randomly repeats the samples from the tail classes, while the latter randomly discards samples from the head classes. Recently however, various other forms of

2. **Cost-sensitive Learning:** Cost-sensitive learning seeks to re-balance classes by adjusting loss values for different classes during training. Recent studies have developed various cost-sensitive long-tailed learning methods to handle class imbalance, including *class-level re-weighting* and *class-level re-margining.* Class-level re-weighting directly uses label frequencies of training samples for loss re-weighting. Class-level re-margining seeks to handle class imbalance by adjusting the minimal margin (i.e., distance), between the learned features and the model classifier, for different classes [1].

3. **Logit Adjustment:** Logit adjustment is a technique that shifts the model logits based on label frequencies. Recently, Menon et al. at Google published a study called "Long-Tail Learning via Logit Adjustment" which proved that logit adjustment is Fisher consistent to minimize the average per-class error.

Information Augmentation

Information augmentation based methods seek to introduce additional information into model training, so that the model performance can be improved in long-tailed learning. There are two kinds of methods in this method type: transfer learning and data augmentation [1].

1. **Transfer Learning:** Transfer learning seeks to transfer the knowledge from a source domain (e.g., datasets, tasks or classes) to enhance model training on a target domain. In deep long-tailed learning, there are four main transfer learning schemes: *head-to-tail knowledge transfer*, *model pre-training*, *knowledge distillation*, and *self-training* [1].

2. **Data Augmentation:** Data augmentation is essentially a set of techniques used to create more instances of training data from the existing training data itself, which enhances the size and quality of the datasets for model training. Two groups of data augmentation methods exist in the field of Deep Long-Tailed Learning including *transfer-based augmentation* and *conventional (non-transfer) augmentation.*

Module Improvement

1. **Representation Learning:** Existing representation learning methods for long-tailed learning are based on four main paradigms, i.e., metric learning, sequential training, prototype learning, and transfer learning. Metric learning aims at designing task-specific distance metrics for establishing similarity or dissimilarity between objects. An example of sequential learning is Hierarchical feature learning (HFL), which hierarchically clusters objects into visually similar groups, forming a hierarchical cluster tree. In this study, Ouyang et al. create such a cluster tree in which the model in the original node is pre-trained on ImageNet-1K; the model in each child node inherits the model parameters from its parent node and is then fine-tuned based on samples in the cluster node. In this way, the knowledge from the groups with massive classes is gradually transferred to their sub-groups with fewer classes. Prototype learning based methods seek to learn class-specific feature prototypes to enhance long-tailed learning performance. Transfer-learning based long-tailed methods that improve representation learning include SSP, LEAP, and unsupervised discovery (UD)[1][3].

2. **Classifier Design:** Different designs of classifiers can also be helpful in addressing long-tailed problems. Various different designs have been explored in many studies, depending on their use-case. A few examples are RTC (Realistic Taxonomic Classifier), Casual classifier, and GIST classifier.

3. **Decoupled Training:** Decoupled training decouples the learning procedure into representation learning and classifier training. In decoupled training, the main observations are twofold: (1) instance-balanced sampling is surprisingly the best strategy for representation learning; (2) the devil is in classification: re-adjusting the classifier leads to significant performance improvement in long-tailed recognition [1].

4. **Ensemble Learning:** Ensemble learning based methods strategically generate and combine multiple network modules (namely, multiple experts) to solve long-tailed visual learning problems [1]. An illustration of existing ensemble-based long-tailed learning methods is shown below:

Another interesting contribution of this survey is the proposition of a new metric, to handle long-tailed learning problems: **Relative Accuracy**.

Given that the main goal of long-tailed learning studies is to allow for better performance of models on handling class imbalances, a new metric has been proposed that attempts to better represent the model's ability to handle class imbalances. A common evaluation protocol is to directly use the top-1 test accuracy to judge how well long-tailed methods perform, and which handles class imbalances better. The problem here is that such metrics cannot accurately reflect the relative superiority of these methods, because the top-1 accuracy could be influenced by factors other than class imbalances. The survey introduced the novel metric Relative Accuracy to alleviate the influence of unnecessary factors in long-tailed learning.

Credit and References:

[1] Deep Long-Tailed Learning: A Survey by Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. https://arxiv.org/pdf/2110.04596.pdf

[2] Long-Tail Learning via Logit Adjustment by Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. https://arxiv.org/pdf/2007.07314.pdf

https://openaccess.thecvf.com/content_cvpr_2016/papers/Ouyang_Factors_in_Finetuning_CVPR_2016_paper.pdf

Deep Learning Machine Learning Class Imbalance