Learning from small samples

Kocaman, V.

Chapter 6

Conclusions

In conclusion, this thesis has delved into the topic of learning from small samples in machine learning and provided a comprehensive overview of the various approaches that have been proposed to effectively tackle this problem.

The third chapter presented an overview of the various techniques used in overcoming small data problems, including data selection and preprocessing, incorporating prior knowledge, ensemble methods, transfer learning, regularization techniques, and synthetic data generation, among others.

The approaches discussed in this chapter covered a wide range of methods, including data selection and preprocessing, incorporation of domain, prior and context knowledge, ensemble methods, transfer learning, parameter initialization, loss function reformulation, regularization techniques, data augmentation, synthetic data generation, problem reduction, optimization techniques, using physics-informed neural networks, unsupervised learning techniques, semi-supervised learning, self-supervised learning, zero-shot, one-shot and few-shot learning, meta learning, harnessing model uncertainty, active learning, self-learning, multi-task learning, symbolic learning, hierarchical learning, knowledge distillation based learning, and dealing with imbalanced data.

The forth chapter focused on the impact of batch normalization in learning from small samples, particularly in the context of imbalanced datasets. To simulate such scenarios, we artificially generate skewness (99% vs. 1%) for certain plant types out of the PlantVillage dataset as a basis for classification of scarce visual cues through transfer learning. By randomly and unevenly picking healthy and unhealthy samples from certain plant types to form a training set, we consider a base experiment as fine-tuning ResNet34 and VGG19 architectures and then testing the model performance on a balanced dataset of healthy and unhealthy images. We empirically observe that the initial F1 test score jumps from 0.29 to 0.95 for the minority class upon adding a final Batch Normalization (BN) layer just before the output layer in VGG19.

In order to find if there is a certain ratio in which the impact is maximized, we experimented with various levels of imbalance ratios and conditions, and observed that the impact of final BN on highly imbalanced settings is the most obvious when the ratio of minority class to the majority is less than 10%; above that almost no impact. As expected, the impact of the final BN layer is more obvious on minority class than it is on majority class, albeit the level of impact with respect to the imbalance ratio is almost same, and levels off around 10%.

We also experimented if the batch size would also be an important parameter for the minority class test accuracy when the final BN is added and found out that the highest score is gained when the batch size is around 64, whereas the accuracy drops afterwards with larger batches. Additionally, we empirically demonstrated that the final BN layer could still be eliminated in inference without compromising the attained performance gain.

Our calibration experiments show that a network has much lower ECE, i.e. has a more ideal confidence relative to its own accuracy. In effect, a final BN-layered network is not 'over-confident' and as a result generalizes and performs better on unseen datasets.

As a result, we empirically illustrated that adding a batch normalization layer before the softmax output layer significantly reduced the training time and improved the test error for minority classes, resulting in a more than three-fold performance boost in some configurations.

The fifth chapter explored the role of self-supervised learning as an augmentation policy in learning from small samples. The study showed that using salient image segmentation in self-supervised learning improved the representations learned, especially in the context of the downstream task of image segmentation.

More specifically, this chapter investigated the impact of the Global Contrast based Salient Region Detection (SGD) algorithm on self-supervised learning. I demonstrated the potential of SGD as a powerful image augmentation technique for image segmentation tasks in self-supervised learning. The implementation of SGD into SSL pretraining routines was achieved through a simple manipulation called offline augmentation with hashing. The experiments carried out showed that using SGD as an augmentation policy in SSL generates better representations for image segmentation tasks.

The results also indicated that SSL with SGD-based augmentation performed well with low resolution images, but this needs further investigation. The results showed that different SSL methods perform differently based on the augmentation policy used and that the impact of SGD also varies in different settings. An unexpected observation was the worse results obtained with SGD applied to high-resolution images compared to low-resolution ones. The results of this study highlighted the importance of the augmentation technique, type of downstream task, and image resolution in determining the success of a SSL method. These findings can be used to guide future studies on self-supervised learning and the application of SGD in this field.

One promising avenue for further investigation in the area of self-supervised learning pretraining would be the integration of offline augmentation with hashing, which could yield valuable insights into the performance of unsupervised and zero-shot segmentation algorithms, as well as salient object detection techniques. Additionally, our results highlight the potential benefits of exploring other types of augmentation policies in the pretraining process, and we encourage researchers to continue exploring these and other approaches to improve the performance and versatility of self-supervised learning models in a wide range of applications.

In summary, this thesis highlights the importance of effectively learning from small samples in machine learning and the various approaches that can be used to tackle this problem. The results of this study demonstrate the potential of batch normalization and self-supervised learning in improving the performance of models trained on small datasets. These findings have important implications for researchers and practitioners working in the field of machine learning, and open up avenues for further exploration and innovation in the area of learning from small samples.