Transfer learning

Mohammed Fayiz, Juntao Jiang, Marfa Zakirova, Zuhaib Khaki, Nazila Eshghi
Tempora pod, NMA 2022

Most often, machine learning algorithms assume that training and test samples contain the same features and belong to the same distribution. However, it happens that for a specific task it is necessary to build a complex model, but there is not enough data for its correct training. Then the application of transfer learning is used. In our project, we consider the task of classifying images on datasets with different numbers of classes.

For implementation, we use the Resnet-18 network pre-trained on ImageNet. Afterwards we tune it for applying on target domain, $CIFAR - 10$; so, the model also learns the distribution to labels provided with this dataset. We change the number of classes in pre-trained model's last layer before the softmax, freeze the remaining weights and gradually improve the model: layer by layer, from the end. The result is exhaustive: when defrosting only two layers, the accuracy reaches 97%.

Then we observe catastrophic forgetting problem. It occurs when a neural network loses the information learned in a previous task after training on subsequent tasks.

We train the previously obtained model on CIFAR-10 subclasses and evaluate the accuracy of the final model. Step 1: training and testing of the 1st and 2nd classes. Step 2: training on 3rd and 4th, checking 1,2,3,4. And so on. As predicted, the accuracy is falling.

In this paper, we give an example of qualitative models' large learning opportunities on a small number of data. Of course, it is worth considering the consequences of forgetting, which motivates to study specialized models to clarify this drawback.